NEW YORK EXPERIMENTS WITH NEW-TYPE MODERN LANGUAGE TESTS

BEN D. WOOD

PUBLICATIONS OF THE
AMERICAN AND CANADIAN COMMITTEES
ON MODERN LANGUAGES

BISHOP'S UNIVERSITY
LIBRARY

NEW YORK EXPERIMENTS
WITH
NEW-TYPE MODERN LANGUAGE TESTS

Publications of the American and Canadian Committees on Modern Languages

NEW YORK EXPERIMENTS WITH NEW-TYPE MODERN LANGUAGE TESTS

By Ben D. Wood

A LABORATORY STUDY OF THE READING OF MODERN FOREIGN LANGUAGES

By G. T. Buswell

ENROLMENT IN MODERN FOREIGN LANGUAGES IN SECONDARY SCHOOLS AND COLLEGES IN THE UNITED STATES

By Andrew T. Wylie

A GRADED SPANISH WORD BOOK*

Compiled by Milton A. Buchanan

*Published by The University of Toronto Press, Toronto, Canada

New York Experiments with New-Type Modern Language Tests

INCLUDING

A SURVEY OF MODERN LANGUAGE ACHIEVEMENT IN THE JUNIOR HIGH SCHOOLS OF NEW YORK CITY, JUNE, 1925;

THE REGENTS EXPERIMENT OF JUNE, 1925, WITH NEW-TYPE TESTS IN FRENCH, GERMAN, SPANISH, AND PHYSICS
WITH A FOREWORD BY THE COMMISSIONER OF EDUCATION OF NEW YORK STATE

AND

A SECOND SURVEY OF MODERN LANGUAGE ACHIEVEMENT IN THE JUNIOR HIGH SCHOOLS OF NEW YORK CITY JUNE, 1926

BY

BEN D. WOOD

DIRECTOR OF THE BUREAU OF COLLEGIATE EDUCATIONAL RESEARCH
COLUMBIA COLLEGE

New York
THE MACMILLAN COMPANY
1927

All rights reserved

COPYRIGHT, 1927,
BY THE MACMILLAN COMPANY.

Set up and electrotyped. Published August, 1927.

PRINTED IN THE UNITED STATES OF AMERICA
BY BERWICK & SMITH CO.

FOREWORD

This monograph is the first of a series to be issued by the Modern Foreign Language Study with the Canadian Committee on Modern Languages, under the auspices of the American Council on Education. These organizations, supported by a grant from the Carnegie Corporation of New York, have undertaken an investigation of the teaching and study of the modern foreign languages and the development of means for their improvement, particularly at the levels of instruction in secondary school and the earlier years in college.

The present work consists of three studies, all based on the application of objective language tests. These tests have been constructed according to techniques tried out and validated during the last four years by the Columbia College Research Bureau for testing achievement in the modern languages.

The first paper describes the results of applying these tests to all students of French and Spanish in the junior high schools of New York City at the final examinations in June, 1925. The author analyzes the tests themselves and presents the results indicated by their administration.

The second study resulted from the administration of similar tests in French, Spanish, German, and Physics, in the June, 1925 examinations of the Regents of the State of New York, when the examination period was divided equally between new-type and old-type tests. Here, also, a careful analysis of the examinations accompanies a presentation of the results shown by correlations between the returns from the new-type and the old-type tests.

The third study relates to a second survey of achievement in French and Spanish in the junior high schools of New York City. This survey was carried out in June, 1926, one year after the survey on which the first study was based, and the report of its results, published as Part Three in this volume, is a logical continuation of the first study.

Professor Wood's three studies open a field of inquiry of great importance and indicate results which are fundamental for modern language testing. The total number of returns involved — for the Regents experiment 31,025, and for the junior high school surveys approximately 50,000 — dwarfs by comparison every previous experiment with new-type tests in the modern languages. Furthermore, the field of administration offered unusually favorable features, presenting in each case groups of schools in a unified system under highly centralized modern language supervision, and, in the

case of the Regents experiment, employing an examination routine long in use. All this made for pupil material as homogeneous as would be possible with so large a number of American school children, and for conditions of administration as nearly identical as our educational groups can supply.

Under these circumstances the possibilities with new-type tests have been given a stern tryout. They have been used to sound the situation as regards achievement and placement in the junior high school and, together with the old-type tests, to measure the same features in all the college-preparing schools of the State. Of the greatest importance is the opportunity which the author has had for a searching analysis of the old and the new types of tests and a comparative study of their efficiency as tools of measurement.

It is hoped that modern language teachers and school administrators, as well as experts in the field of educational theory, will give the reports on these experiments careful examination. It has been necessary in certain parts of the reports to use a terminology which may seem to some modern language teachers unfamiliar and technical. However, teachers of the foreign languages are coming to recognize the importance of recent educational research for their work and are familiarizing themselves in ever-increasing numbers with the terminology which the use of statistical methods involves. To those readers who do not readily comprehend the more intricate discussions of correlations, probable errors, deviations, and the like, the following reports offer nevertheless in non-technical language a wealth of significant detail on a subject of paramount importance for the organization of modern language teaching.

The first two studies were carried through quite independently of each other and the reports were prepared without thought of joint issue. This accounts for numerous repetitions in the text, particularly in explanation of test techniques, forms of tables, graphs, etc. The reader will, however, become convinced that the studies belong under one cover, since the results which they offer are secured by the same type of test and indicate the same weaknesses in the present system of grading and classifying modern language pupils. To understand the possibilities of tests of this character it is very helpful to be able to traverse the results derived from two fields where the new-type tests have been applied under different conditions. The third study was originally prepared for separate publication, but since it became available before this volume got through the press, it is included because of its close relations to the first study.

The Modern Foreign Language Study
The Canadian Committee on Modern Languages

CONTENTS

PART ONE

SURVEY OF MODERN LANGUAGE ACHIEVEMENT IN THE JUNIOR HIGH SCHOOLS OF NEW YORK CITY

I. Introduction — PAGE

The tests . . . 3
Method of administering and scoring the tests . . . 4
Application of the test results . . . 7

II. The Test Results: Class Norms

Median and quartile scores by classes . . . 9
Individual differences displayed . . . 9
Student eliminations . . . 12
Distributions of scores by classes . . . 15
Students reclassified on uniform basis . . . 15
Fallacy of the "time-serving" basis for classification . . . 16
Need for comparable measurements in junior and senior high schools . . . 18
Equivalence of tests determined experimentally . . . 18
Distributions of scores by parts . . . 19
Overlapping of classes displayed graphically . . . 19
Definition of a percentile curve . . . 19
Spanish classes less homogeneous than French classes . . . 27
Standards in individual schools . . . 27

III. Reliability and Validity of Tests . . . 39
Reliability of Tests . . . 39
Definition of reliability . . . 39
Reliability of old-type examinations . . . 39
Reliability of new-type tests . . . 40
Desirability of 500-question tests . . . 41
Reliability of each Part of the new-type tests . . . 41
Standard and probable errors of estimate . . . 42
Validity of Tests . . . 43
Intercorrelations of parts of new-type tests . . . 43

IV. Analysis of Individual Questions in New-Type Tests . . . 45
Difficulties of Individual Questions . . . 45
Individual questions experimentally validated . . . 45
Order of difficulty correlations . . . 46
Reproductions of the tests . . . 47
New-type French test . . . 47
New-type Spanish test . . . 67
Distributions of Henmon and Wood frequencies of words in French vocabulary test . . . 86
Intercorrelations of Henmon and Wood frequencies and empirical difficulties of words in French vocabulary test . . . 86

PAGE

Validity of Individual Questions 87
Definition of a "good" question 87
Possible causes of "inverted" questions 88
"Inversions" should be detected and eliminated 88
Proportions of "good" and "bad" questions in the tests 90
Analysis of some "bad" questions 90
Graphs of some "good" and "bad" questions 91

V. Summary and Recommendations 94
Importance and difficulty of constructing good tests 94
The essential purpose of good examinations 94
Criteria of good examinations 95
Some limitations of the new-type tests 96
Some objections considered 96
Reconstruction of curriculum dependent upon use of good tests . . . 98
Learning students prerequisite to *teaching* them 99
The "trial-period" in educational guidance 99
Reasons for opposition to new-type tests 100
Constructive usefulness the only justification for tests and examinations 102

PART TWO

THE REGENTS EXPERIMENT WITH NEW-TYPE TESTS IN FRENCH, SPANISH, GERMAN, AND PHYSICS, JUNE 1925.

I. Introduction 107
Plan of the experiment 108
The examinations 109
Only one new-type examination in each modern language for all year-classes in that language; comparability of measures 110
Scope of the experiment 111

II. Reliability and Validity of the Examinations 113
Reliability of Old- and New-Type Examinations 113
Reliability of old-type examinations 113
Reliability of new-type examinations 115
Reliability coefficients of alienation 117
Standard errors of estimate 118
Probable errors of estimate 120
Probable errors of placement 121
Validity of New-Type Examinations 122
Correlations of new- with old-type examinations 122
New-type tests more valid for fourth- and third- than for second-year classes 124
Correlations of parts of new-type tests with old-type examinations . 124
Inter-correlations of parts of new-type tests 125

III. School Differences: Validity of Old-Type Examination Results in Individual Schools and Classes of Schools 129
New-type results used as criteria for measuring validity of old-type examination ratings 129
Validity of ratings from individual schools extremely variable 129

PAGE

Validity of school ratings increases consistently with size of schools . 129
Validity of ratings in such high schools as Morris, Boys, Albany, etc., nearly twice as great as in academies 131
Validity of Regents review ratings varies as much from school to school as validity of school ratings varies, and is only slightly greater, on the whole, than validity of school ratings 132
Reasons for failure of Regents review ratings to correct inaccuracies of school ratings on old-type Regents examinations 133
Not all papers are reviewed 134
Some reviewing is misplaced 134
Review ratings are not independent 134
Unclaimed papers very rarely reviewed 135
Fundamental weakness inherent in old-type subjective form of examinations 135
Classes in large schools are more homogeneous and achieve higher average scores, than in small schools 136
Proportions of students of French in various classes of schools . . . 137
School differences with respect to achievement in vocabulary, reading and grammar 140
Homogeneity of classes 140
Unevenness of achievement 142
Intercorrelations of new-type parts for individual schools 142

IV. Overlapping of Classes and Variable School Standards not Corrected by Regents Examinations 146
Distributions of New-Type Examination Scores 146
Distributions of total scores on new-type tests in French, Spanish, German and Physics for each year-class and for "passed" and "failed" Groups (Tables 17–20) 146
Distributions of scores on each Part of new-type tests, and on Parts I and II combined (Tables 21–24) 150
Overlapping of Classes 154
Means and sigmas of class distributions presented graphically . . . 154
Fallacy of customary assumption underlying use of separate examinations for year-classes 155
"School overlapping" distinguished from "Regents overlapping" . 156
School overlapping displayed by percentile graphs 156
How to read a percentile graph 157
One-third of students in French classes misplaced one year or more . 157
Sixty per cent misplaced one semester or more 159
Regents overlapping 162
Percentile graphs for students "passed" by Regents review ratings . 162
Regents overlapping nearly equal to school overlapping 167
Percentile graphs for students "failed" by Regents review ratings . 167
Percentile graph for new-type Physics test scores of total, passed and failed groups 171
Overlapping of classes displayed by frequency polygons 173
Forms of distributions are evidence against old-type examinations . 181
Regents examinations should be strengthened, not abandoned or curtailed because of their weaknesses 183
Variable Standards of Schools Not Equalized by Regents Examinations . 183

PAGE

Variable standards displayed by comparing new- and old-type average scores of French II students in individual schools and groups of schools 185
Variable standards displayed by Physics test results 185
Variable standards displayed by comparison of new-type average scores of second- and third-year French classes in individual schools and groups of schools 189
Variability of Regents standards only one aspect of fundamental inaccuracy of subjective examinations 193
Unevenness of achievement in individual schools with respect to vocabulary, reading and grammar 197

V. Internal Structure of Old- and New-Type Modern Language Examinations 198
Methods of construction 198
Reproductions of old-type examinations 199
French II old-type examination (2 years) 200
French III old-type examination (3 years) 202
French IV old-type examination (4 years) 204
Physics old-type examination 206
Method of constructing new-type tests 207
Advantages of *One* examination for all year-classes 208
Reproductions of new-type Regents examinations 209
New-type French test 210
New-type Spanish test 229
New-type German test 250
New-type Physics test 270
Size of vocabulary sampling 283
Vocabulary sampling of ninety-minute new-type test about twice as large as that of three-hour old-type examination 283
Character of vocabulary sampling 284
Henmon and Wood French word frequency distributions 284
New-type superior to old-type vocabulary sampling according to Henmon and Wood French word lists 286
Old-type second-, third-, and fourth-year examinations nearly equal as to vocabulary difficulty, and probably also as to grammar and syntax 290
Proportions of high- and low-frequency words in Regents and College Entrance Examination Board old-type examinations, and in Regents and junior high school new type tests 292
Empirical difficulties of individual questions in new-type tests . . . 296
Correlation of empirical difficulties of words with Henmon and Wood frequencies 296
Reliability of orders of difficulty 297
Validity of individual new-type questions 299
Need for empirical verification of measurement values of individual questions 300
Validity of individual old-type questions 303
General weaknesses of translation and free composition questions . . 305
Fallacy of substituting oral credit for free composition questions . . 306
Special weaknesses of free composition question 306
Irrelevant activities in old- and new-type examinations 307

PAGE

VI. Costs of Old- and New-Type Examinations 311

VII. Conclusions and Recommendations 313
History and methods of Regents examinations 313
The Regents system sound in principle 315
Destructive criticisms of Regents system not justified by facts . . . 315
Duties of examining agencies include research and information services . 316
Time-serving conception of achievement is indefensible 317
Summary 318

PART THREE

SECOND SURVEY OF MODERN LANGUAGE ACHIEVEMENT IN THE JUNIOR HIGH SCHOOLS OF NEW YORK CITY, JUNE, 1926

Introduction 323
Numbers of students tested in 1925 and in 1926 323
Equivalence of Forms A and B 325
Influence of homogeneity of classes on achievement 327
Constancy of progress of individuals and of classes 327
Extent of reclassification in 1925–1926 331
Variable standards of schools 332
Graphic analysis of progress of classes in individual schools in relation to initial achievement and to initial homogeneity 334
Summary 338

LIST OF TABLES

PART ONE

TABLE		PAGE
1.	Numbers of students tested in each class and their quartile and median scores	9
2.	Gross apparent eliminations from junior high school modern language classes	12
3.	Distributions of scores on French test	13
4.	Distributions of scores on Spanish test	14
5.	Distribution of scores on each Part of the French and Spanish tests	20
6.	Reliability coefficients of French and Spanish tests	40
7.	Reliability coefficients of the Parts of the French and Spanish tests	41
8.	Intercorrelation of the Parts of the French and Spanish tests	44
9.	Difficulties of individual questions in the French and Spanish tests	46
10.	Distribution of French words in Part I of the French test according to Henmon and Wood frequencies	87
11.	Distributions of individual questions in French and Spanish tests according to degree of validity	89

PART TWO

12.	Numbers of students taking the new-type Regents examinations in French, Spanish, German and Physics	112
13.	Reliability coefficients of old-type Regents examinations in French and Spanish	115
14.	Reliability coefficients of new-type Regents examinations by Parts and wholes	116
15.	Reliability coefficients of alienation of old- and new-type Regents examinations	118
16.	Standard errors of estimate of new-type Regents examinations	120
17.	Probable errors of estimate of new-type Regents examinations	121
18.	Validity coefficients and standard errors of estimate of new-type Regents examinations	123
19.	Validity of Parts of Regents new-type modern language tests	125
20.	Intercorrelation of Parts of Regents new-type modern language tests, with sigmas and standard errors of estimate	126
21.	Intercorrelations between new-type scores, and school ratings and review ratings of old-type Regents examinations in French for individual schools and classes of schools	130
21*a*.	Standard errors of estimate for six size-groups of schools	132
22.	Intercorrelations between new-type scores, and school ratings and review ratings of old-type Regents examinations in Spanish for individual schools and classes of schools	138
23.	Intercorrelations between new-type scores, and school ratings and review ratings of old-type Regents examinations in German for individual schools and classes of schools	139
24.	Intercorrelations between new-type scores, and school ratings and review ratings of old-type Regents examinations in Physics for individual schools and classes of schools	140

TABLE PAGE

25. Intercorrelations of vocabulary, reading and grammar Parts of new-type French examination for individual schools 141
26. Intercorrelations of vocabulary, reading and grammar Parts of new-type Spanish examination for individual schools 143
27. Intercorrelations of vocabulary, reading and grammar Parts of new-type German examination for individual schools 144
28. Distributions of total scores on new-type Regents examination in French, of "passed" and "failed" groups 147
29. Distributions of total scores on new-type Regents examination in Spanish, of "passed" and "failed" groups 148
30. Distributions of total scores on new-type Regents examination in German, of "passed" and "failed" groups 149
31. Distributions of total scores on new-type Regents examination in Physics, of "passed" and "failed" groups 150
32. Distributions of scores on each Part of the new-type Regents examination in French by year-classes 151
33. Distributions of scores on each Part of the new-type Regents examination in Spanish by year-classes 152
34. Distributions of scores on each Part of the new-type Regents examination in German by year-classes 153
35. Distributions of combined scores on the vocabulary and reading Parts of the new-type Regents examinations in French, Spanish and German . . . 154
36. Overlapping of modern language classes 171
37. The Henmon and Wood French word lists 285
38. Distributions of Henmon and Wood frequencies of French root words in the old-type Regents and College Entrance Examination Board French examinations 286
39. Distributions of Henmon and Wood frequencies of French root words in the new-type junior high school and Regents examinations in French of June 1925 287
40. Distributions of empirical difficulties of individual new-type questions . . 297
41. Reliabilities of orders of difficulty of questions in new-type Regents examinations in French 298
42. Distribution of individual questions in new-type Regents examinations according to degree of validity 300
43. Correlations of new-type examination scores with the translation and free composition questions in the old-type Regents examinations in French and Spanish 304
44. Numbers of times the same French root word is used in old- and in new-type Regents examinations in French 308
45. Relative costs of scoring old- and new-type examinations 312

PART THREE

46. Numbers of New York City junior high school students taking the American Council Beta Tests in French and Spanish in 1925 and 1926 323
47. Equivalence of Forms A and B of American Council Beta Tests; showing medians and quartiles of each class in French and Spanish 326
48. Constancy of Progress in Modern Language Work. Correlations between scores on Form A taken in June 1925, and scores on Form B, taken in June 1926 328

TABLE PAGE

49. Constancy of Progress in modern language work. Correlations between 1925 medians and 1926 medians of individual school classes made up of the same individual students 329
50. Numbers and proportions of students who scored lower on Form B in June 1926 than on Form A in June 1925 330
51. Comparison of variabilities of classes in individual schools. 331
52. Comparison of medians of classes in individual schools 333

LIST OF CHARTS

PART ONE

CHART		PAGE
1.	Graph of medians and interquartile ranges of scores on the French and Spanish tests of each class	10
2.	Percentile curves of scores of the 8A, 8B, 9A and 9B French classes	22
3.	Percentile curves of scores of the RB, RC and RD French classes	24
4.	Percentile curves of scores of the 8A, 8B, 9A and 9B Spanish classes	28
5.	Percentile curves of scores of the RB, RC and RD Spanish classes	30
6.	Graphs of distributions of French test scores of 8A, 8B, 9A and 9B classes	32
7.	Graphs of distributions of French test scores of RB, RC and RD classes	32
8.	Graphs of distributions of Spanish test scores of 8A, 8B, 9A and 9B classes	33
9.	Graphs of distributions of Spanish test scores of RB, RC and RD classes	33
10.	Graphic comparison of median scores of each "normal" French class in each school	34
11.	Graphic comparison of median scores of each "rapid" French class in each school	36
12.	Graphic comparison of median scores of each "normal" Spanish class in each school	37
13.	Graphic comparison of median scores of each "rapid" Spanish class in each school	38
14.	Graphs illustrating the validity of some of the best individual questions in the new-type French and Spanish tests	92
15.	Graphs illustrating the lack of validity of some of the least effective questions in the new-type French and Spanish tests	93

PART TWO

16.	Graphs of means and sigma ranges of scores on new-type modern language examinations	155
17.	Percentile curves of new-type French test scores of 2nd, 3rd and 4th year classes	158
18.	Percentile curves of new-type Spanish test scores of 2nd, 3rd and 4th year classes	160
19.	Percentile curves of new-type German test scores of 2nd, 3rd and 4th year classes	161
20.	Percentile curves of new-type French test scores of "passed" students by year classes	163
21.	Percentile curves of new-type Spanish test scores of "passed" students by year classes	164
22.	Percentile curves of new-type German test scores of "passed" students by year classes	165
23.	Percentile curves of new-type French test scores of "failed" students by year classes	168
24.	Percentile curves of new-type Spanish test scores of "failed" students by year classes	169
25.	Percentile curves of new-type German test scores of "failed" students by year classes	170

CHART PAGE

26. Percentile curves of new-type Physics test scores of the entire group, of "passed" and of "failed" students 172
27. Graphs of distributions of new-type Regents French examination scores of "passed" and "failed" students in each year class 174
28. Graphs of distributions of new-type Regents Spanish examination scores of "passed" and "failed" students in each year class 175
29. Graphs of distributions of new-type Regents German examination scores of "passed" and "failed" students in each year class 176
30. Graphs of distributions of new-type scores of total, "passed" and "failed" groups of each French class 177
31. Graphs of distributions of new-type scores of total, "passed" and "failed" groups of each Spanish class 179
32. Graphs of distributions of new-type scores of total, "passed" and "failed" groups of each German class 180
33. Graphs of distributions of new-type Physics test scores of total, "passed" and "failed" groups 181
34. Graphs of distributions of French, Spanish, German and Physics old-type examination grades by classes 182
35. Variability of Regents standards in French II 184
36. Variability of Regents standards in Physics 186
37. Graphic comparison of new-type median scores of French II and of French III classes in individual schools and groups of schools 187
38. Graphic comparison of new-type median scores of "passed" students in French II and in French III classes in individual schools and groups of schools 188
39. Graphic comparison of new-type median scores of "failed" students in French II and in French III classes in individual schools and groups of schools . 190
40. Graphic comparison of new-type medians and interquartile ranges of "passed" and "failed" students in French II in individual schools 192
41. Graphic comparison of new-type medians and interquartile ranges of "passed" and "failed" students in French III in individual schools 194
42. Graphic comparison of new-type medians and interquartile ranges of "passed" and "failed" students in Spanish II in individual schools 195
43. Graphs showing unevenness of achievement of modern language classes in individual schools with respect to achievement in vocabulary, silent reading and grammar 196
44. Graph showing distributions according to Henmon frequencies of different root words in old- and new-type French tests 289
45. Graph showing proportions of different root words in old- and new-type French examinations which are in very high or very low Henmon frequency groups 291
46. Graph showing proportions of different root words in old- and new-type French examinations which are common to indicated numbers of French text-books 295
47. Graphs illustrating varying degrees of validity of individual new-type questions 302

PART THREE

48. Graphs of 1925 and 1926 French class medians (from Table 47), showing equivalence of Forms A and B of the American Council Beta French Test 324
49. Graphs of 1925 and 1926 Spanish class medians, showing equivalence of the American Council Beta Spanish Test 325

CHART PAGE

50. Graphs of medians and interquartile ranges of American Council Beta French Test scores of identical groups of students who took Form A in June 1925 as 8A students and Form B in June 1926 as 9A students 335

51. Graphs of medians and interquartile ranges of American Council Beta French Test scores of identical groups of students who took Form A in June 1925 as 8B students and Form B in June 1926 as 9B students 336

52. Graphs of medians and interquartile ranges of American Council Beta French Test scores of identical groups of students who took Form A in June 1925 as RB students and Form B in June 1926 as RD students 337

AUTHOR'S ACKNOWLEDGMENTS

This volume is published by the Modern Foreign Language Study, and is the result of the coöperation of many individuals and institutions.

The construction of the tests used in the junior high schools was made possible by three grants from The Commonwealth Fund, and by the coöperation of the Bureau of Collegiate Educational Research in Columbia College; the survey was made at the instance of Mr. Jacob Greenberg, Director of Foreign Languages in the Junior High Schools of New York City; the statistical analysis of the returns has been made possible by grants from The Commonwealth Fund, The Modern Foreign Language Study, and the Bureau of Collegiate Educational Research in Columbia College.

The writer owes special thanks to Mr. Greenberg for his active and disinterested coöperation in this and in other similar research projects during the last four years. Acknowledgment is made of the valuable assistance rendered by Miss Blanche Allain, Teacher of French in the junior high schools of New York City; by Professor Frank C. Callcott and Mr. R. H. Williams, co-authors of the Spanish Test; and by many principals and teachers in the junior and senior high schools in New York City.

The main credit for the Regents experiment belongs to the New York State Department of Education, particularly to Dr. James Sullivan, who initiated it, and to Dr. Avery W. Skinner, who was in immediate charge of its administration. The construction of the new-type examinations and the writer's contribution to the experiment were made possible by a grant from The Commonwealth Fund and by the provisions for research made by the office of the Dean of Columbia College. The conduct of the experiment was made possible by a grant from the Carnegie Corporation to the Regents of the University of the State of New York. The analysis of the results and the writing of the report were financed by The Commonwealth Fund and The Modern Foreign Language Study.

The experiment would have been impossible without the work of the subject-matter specialists who, with technical assistance from the writer, constructed the tests; the late lamented Professor A. A. Méras, Miss Suzanne Roth, Professors C. M. Purin, Frank Callcott, and Hermon W. Farwell. Their contributions, involving in some cases more than three years of experimental work under trying conditions, can hardly be overestimated. Thanks are extended to Dr. William R. Price for his careful work in checking and correcting the manuscripts of the three modern

language tests and for other valuable help. It is impossible to make detailed acknowledgments to the host of principals and teachers, scattered from New York to California, whose coöperation during several years made available the experimental evidence on the preliminary questions and question-forms which guided us in the construction of the tests.

The publication of these reports under one cover has been financed by The Modern Foreign Language Study. Thanks are due and are heartily made to Robert Herndon Fife, Professor of Germanic Languages in Columbia University, and to Professors V. A. C. Henmon, Algernon Coleman, and C. M. Purin for their unfailing coöperation and generous advice. Their careful editorial assistance makes this volume a contribution of The Modern Foreign Language Study in a very real sense.

The author's main debt of gratitude is to Dean Herbert E. Hawkes of Columbia College, and to Professor J. J. Coss, without whose encouragement and guidance this work would not have been possible. Professors E. L. Thorndike and A. L. Jones, and Mr. Frank D. Fackenthal have shared with Dean Hawkes and Professor Coss the task of directing the writer's part in this enterprise, and acknowledgment is gratefully made of their indispensable help.

The writer is especially indebted to Mary Eleanor Wood, Helen J. Green and Esta S. Flood for able assistance in the technical conduct of the experiment and in the analysis of the results and writing of this report. Mrs. Flood's contributions to the analysis of the data and her careful work in the preparation of the manuscript of the report can hardly be adequately acknowledged. It can only be said that without her devoted and intelligent assistance this report could not have been produced at this time.

Ben D. Wood

Columbia College
April, 1927

PART ONE

SURVEY OF MODERN LANGUAGE ACHIEVEMENT IN THE JUNIOR HIGH SCHOOLS OF NEW YORK CITY

I

INTRODUCTION

At the request of Mr. Jacob Greenberg, Director of Foreign Languages in the Junior High Schools of New York City, I undertook in the spring of 1925 to assist in the construction of tests which should afford objective, reliable, and comparable measures of achievement in French and in Spanish, and which should be adapted to the whole range of achievement in modern language classes in the junior high schools. It was agreed that two equivalent examinations (Forms A and B) should be constructed for each language; that, in accordance with the indications of recent experiments in modern language testing,[1] the examinations should attempt to measure vocabulary, reading-comprehension, and grammar, without attempting to measure the oral-aural abilities and other less measurable values of modern language teaching; and that the examinations should be constructed not solely on the basis of the New York City junior high school syllabi in French and Spanish, but should take advantage of the word-counts [2] which had been made since the production of the syllabi. Fortunately, there were no irreconcilable conflicts between the syllabi and the results of recent investigations on the common fundamentals of vocabulary, grammar, and idioms, so that it was possible to produce tests which were valid not only for the junior high schools of New York City but for elementary modern language classes all over the country.

The French examination (Forms A and B) was constructed in collaboration with Mr. Greenberg and Miss Blanche Allain; and the Spanish examination (Forms A and B) in collaboration with Professor Frank C. Calcott and Mr. R. H. Williams, of the Spanish Department of Columbia University.

The tests. — The tests in French and Spanish are exactly parallel in form, each consisting of three Parts:

Part I is a vocabulary test of 100 foreign language words, each of which is followed by five English words. The student shows his knowledge of the word by selecting the one English word in the five which

[1] See Méras, Roth, and Wood: "A Placement Test in French," *Contributions to Education*, Vol. I, J. C. Bell, Editor, World Book Co., Yonkers, N. Y. 1925.

[2] See e.g. Henmon's *The French Word Book*, University of Wisconsin Bulletin No. 3, September, 1924; Wood's *A Comparative Study of the Vocabularies of Sixteen French Textbooks*, The Modern Language Journal, February, 1927; and Jamieson *A Standardized Vocabulary for Elementary Spanish*, The Modern Language Journal, March, 1924; and the syllabi and vocabulary lists of the junior and senior high schools of New York City.

corresponds in meaning to the French or Spanish word. The vocabulary words were selected mainly from among the most frequently used 2000.

Part II is a reading-comprehension test, consisting of 60 carefully graded statements in French or Spanish, respectively, each of which has five alternative endings. Only one of the endings is coherent or true, and the student shows his understanding of each statement by selecting this one ending.

Part III is a grammar test consisting of 60 or 65 short English sentences or phrases, each of which is followed by a French or Spanish translation which is incomplete. The student shows his knowledge of the grammatical and other points involved by completing the translations.

The selection of these forms of questions is based upon four years of continuous experimenting with many kinds and forms of questions. Their traits will be discussed at length later in this report.

Each test thus consists of 220 or 225 elements, which constitute a sampling of modern language materials much broader and more varied than is possible in the old-type subjective examinations. The students are allowed thirty minutes for each Part, or ninety minutes for the whole test. The broad sampling is made possible by the fact that little time is lost in irrelevant activities, such as writing out translations of whole sentences and paragraphs. Writing out the translation of a sentence in longhand is not the only way in which a student may unambiguously indicate an exact understanding of the sentence, or ability to translate it. When we consider the complete objectivity of the scoring of the responses to this large number of carefully graded questions, it is not surprising that we achieve with examinations of this type results which are two or three times as reliable as those obtained with old-type tests. It is the objectivity of these tests which makes possible a comparison of their measurements. It is hoped that the importance of comparability in educational measurements will be made clear in succeeding sections of this report. Form A of each test is reproduced below on pages 47–85. The method of selecting the items for these tests and of equating the two forms of each has been described in detail in the article already cited on page 3. The method is a standardized procedure which has been used in principle in the construction of nearly all standard and valid examinations in other subject matters. The principles underlying the method have been set forth in a large number of publications. The most convenient source for the modern language teacher is probably the *Manuals of Directions* for the Columbia Research Bureau Tests in Modern Languages recently published by the World Book Company, Yonkers, N. Y.

Administering and scoring the tests. — The modern language work in the junior high schools of New York City normally runs through two

years of time, or four semesters, and is designed to cover the first year of modern language work in the senior high schools. The normal semester classes are designated in the usual way from lowest to highest as 8A, 8B, 9A, and 9B. Students who successfully complete all four semesters are normally eligible to enter second-year classes in the senior high school. Parallel with the four-semester course we have the three-semester or "rapid advancement" course, which covers the same material as the four-semester or normal course, but does it in three semesters. Students are put in the rapid advancement group on the basis of general scholarship, and not on the basis of particular aptitude for modern language work. The three rapid advancement classes are designated from lowest to highest as RB, RC, and RD.

Until the present survey was undertaken it had been customary to give a uniform examination each year to the 9B and RD classes, but not to any other modern language classes in the junior high schools. The main purpose of this uniform examination *at the end* of the junior high school modern language course was to eliminate the unfit, and allow only well-prepared students who gave promise of succeeding in senior high school modern language work to enter second-year classes in the senior high school. Thus the differentiating function of the junior high school was invoked on a uniform city-wide standard mainly at the end of the junior high school modern language course. It was in order to facilitate adjustments and constructive educational guidance *during* the junior high school course, as well as at its *end,* that these tests were designed to cover the whole range of achievement in all four semesters, thus affording accurate and comparable measures and valid city-wide standards for all four normal semester classes and all three rapid advancement classes. This use of a single scale of achievement for all classes represents a forward step in the administration of modern language instruction which is fundamental and of far-reaching importance.

The old custom of giving separate subjective examinations to each semester- or year-class is slowly but surely disappearing, and deservedly so, because it is indefensible. It is part and parcel of the old time-serving conception of achievement, which was based upon ignorance or disregard of the fundamental fact of individual differences. The incongruity of attempting to apologize for an examination system which often *passes* a student in third-year French and on the same day *fails* him in second-year French, while denying admission to third-year French examinations until the full sentence of three years has been served, is too manifest even to the most casual observer.

In June, 1925, the new-type Comprehensive Examination in French for the Junior High Schools of New York City was administered to nearly 19,000 students taking French: 8A, 8B, 9A, 9B, RB, RC, and RD; and

the corresponding Spanish examination to nearly 6,500 students taking corresponding semesters of Spanish in the junior high schools of New York City. The tests were administered and scored by the teachers. Some irregularities occurred, as is always the case when such large numbers of teachers and pupils are concerned; but they were very few in number and of small consequence. In so far as they had any influence they tended to make the tests seem less reliable than they really are. Several conferences were held by Mr. Greenberg with the entire staff of teachers, at which the whole procedure for administering and scoring the tests was explained in detail. In addition, mimeographed sheets with complete directions for administering the tests, and detailed scoring procedure, were distributed.

In order to make the results of this first trial of an objective and comparable scale as reliable as possible, it was planned to check all the scoring in Mr. Greenberg's office. By this procedure it was thought that other irregularities might be detected and corrected. The first paragraph on the mimeographed sheets distributed to the teachers at the conference read as follows:

"In scoring tests of this type, the mechanical method used is of great importance. Since the scoring will be checked in the office of the Director of Foreign Languages, it is imperative that each teacher should use the same system of marking. A scoring key will be furnished to each department head, from which a sufficient number of copies will be made for each teacher in the department. No deviations from the key and no additions to it should be made by any teacher without consulting the department head. All changes in, or additions to, the key should be reported in detail to the office of the Director."

The task of checking the scoring proved to be too great, so that only a sampling of the papers from each school were checked. Nearly all of the mistakes found were mistakes of arithmetic. As was fully expected, several additions had to be made to the key for Part III. The directions to the teachers included the following:

"The score on each Part is the number of correct answers. No answer in any Part is to receive any credit unless it is absolutely correct, complete, and unambiguously indicated or written; but students should not be penalized for poor penmanship, and in Part III legitimate equivalents of the key answers should receive full credit. In Part III absolute correctness must be rigorously enforced: no answer is to receive any credit if it is in *any* respect — spelling, punctuation, capitalization, etc. — deficient or incomplete. Each correct answer in each Part should be marked immediately at its right with a short, straight, neat, horizontal line, in red or blue crayon: the score on each Part is the number of these red or blue lines (not checks!). The scores on the Parts should be entered,

after the counting has been checked, on the cover page in the spaces provided; the total score is the sum of the three Part-scores. It is important that the red or blue marks be written neatly and clearly, in a straight vertical column, so that they may be easily and accurately counted. Check marks should be avoided; they are clumsy and confusing and render the checking of the scoring and of the counting very difficult. A little care and regularity and neatness in marking the correct answers will avoid many serious errors."

Application of the test results. — As soon as each teacher had completed the scoring of her papers, and had entered the scores in her class roll, the papers were sent to the Office of the Director. Distributions of the scores of each class on a city-wide basis were made as rapidly as the papers arrived. On the basis of these distributions the passing and failing grades for each class were established, and tables for converting scores into the usual percentage grades were sent to each department head. If an 8A student made a score above the passing line for 9A or 9B, he was given a passing grade in 9A or 9B; if a student in 9A or 9B secured a score below the passing line of 8B or 8A, he was failed not only in 9B and 9A, but also in 8B and 8A. The adoption by Mr. Greenberg of this method of assigning grades, which will be explained in detail in succeeding pages, is a notable event, because so far as the writer knows, it is the first instance in the history of modern language teaching in which the principle of classification on the basis of actual individual achievement was, at a single stroke, substituted for the "time-serving" basis. The reader should notice that however radical and drastic the scheme may appear, it is in reality nothing but a sane and logical extension of the sound principle already recognized in the organization of rapid advancement classes, and that it tends to a more effective and fuller realization of individual capacities of students than is otherwise possible.

Administrative exigencies in the various schools, and the difficulties attending promotions and demotions by single subject-matters rather than by school-grade, made it impossible to take anything like full advantage of the city-wide reclassification which these tests made possible. Whatever significance this survey may have, however, does not depend upon whether a reclassification was effected in 1925 or in any other one year: its significance lies rather in the fact that it has uncovered a crucial need for sound classification the extent of which was hardly suspected, and at the same time has demonstrated a sure and simple way for securing in exact and comparable units the measurements which are a *sine qua non* for the proper classification and educational guidance of students with respect to modern language work.[1]

[1] The significance of these disclosures is augmented by the confirmatory evidence uncovered by the use of new-type tests in the Regents Examinations in French, Spanish, and German of June, 1925. See particularly pp. 146 to 198 of this volume.

In general terms, this survey shows how over 25,000 young hopefuls in the largest city in the country are starting their modern language careers; it emphasizes the tremendous importance of individual differences, and the stark fallacy of taking calendar-time spent "studying" a modern language or rather attending modern language classes, as a measure of achievement in that language.

NOTE. — In reviewing the data of this survey we cannot escape the suspicion that only a fraction of these 25,000 youngsters really belong in foreign language classes. This suspicion becomes almost a conviction when we compare the foreign language situation in this country with that in European countries.

Foreign language instruction is one of the oldest and most important elements in the *liberal* curriculum in all the countries of western civilization. For many centuries it has been a successful and exceedingly useful part of the education of cultivated people. No part of the liberal curriculum in Europe enjoys (or deserves) a greater prestige; but it has no such place in the primary school curriculum.

In this country we have tried to put foreign language study in the common curriculum and have unjustifiably assumed it would be as successful as it has been in what the Europeans call *secondary* education. We have inherited the prestige of foreign language study without inheriting all the conditions which created that prestige centuries ago, and which have maintained it intact to this day, in Europe. Three of the most important conditions which have made foreign language study successful in Europe deserve special mention because of their sharp contrast with conditions in this country.

In Europe foreign language instruction (*a*) is restricted to highly selected students, who are really competent to learn a foreign language in addition to the native language, and who can reach a general culture level high enough to make mastery of a foreign language useful in adult life activities; (*b*) is given by teachers of a high general culture level, whose mastery of the language itself and whose knowledge of the culture and civilization of the people that use it are above question, and who have had expert pedagogical training; and (*c*) is continued, when once begun, long enough to make possible a genuine mastery of the language itself and a real acquaintance with the culture, civilization and history of the people that use it.

The present report is mainly concerned with the problem of bringing about in America the first condition mentioned above, namely, that of restricting foreign language instruction to students who are really competent to learn more than one language in a way that will be useful and satisfying to them. The ideal is to define foreign language objectives in relation to the capacities, interests, and needs of individual students. Tentative decisions with regard to individual students can be made on the basis of intelligence test results, facility with the native language, and general school standing; but borderline cases can be safely disposed of only by actual trial, with close observation of progress by means of such objective and comparable measuring devices as the tests used in this survey.

Not the least important result of better educational guidance and of earlier elimination from foreign language work of students not fitted for it by original capacities, interests, and needs would be an immediate possibility of raising the general and special culture levels of the foreign-language teaching groups. The indiscriminate foreign language requirement for all kinds of students has so inflated the class rolls that many schools have had to employ substitute teachers, and the general shortage of teachers has prevented any notable raising of the licensing standards. Some one has remarked that if foreign language classes were restricted to the right kinds of students, there would be enough good teachers to go around. The really competent students, having the undivided attention of better teachers, would respond at once to the sense of real achievement, and larger proportions of them would willingly continue foreign language study long enough to reach a genuinely satisfying stage of mastery.

II

THE TEST RESULTS: CLASS NORMS

Median and quartile scores by classes. — Table 1 shows the numbers of students in each class in French and in Spanish, and the median and lower- and upper-quartile scores of each such class.

TABLE 1

NUMBERS OF STUDENTS TESTED IN EACH CLASS IN FRENCH AND SPANISH IN THE JUNIOR HIGH SCHOOLS OF NEW YORK CITY, AND THEIR QUARTILE AND MEDIAN SCORES, JUNE, 1925.

French

	8A	8B	9A	9B	RB	RC	RD
Number . . .	3,736	3,041	2,604	2,417	2,766	2,415	1,898
Lower Quartile .	29.2	47.0	70.5	110.2	40.0	75.5	116.7
Median . . .	39.5	61.0	87.3	130.0	51.2	94.0	140.2
Upper Quartile	50.0	77.7	105.5	148.0	64.3	111.5	160.5

Spanish

	8A	8B	9A	9B	RB	RC	RD
Number . . .	1,486	1,332	658	688	850	683	721
Lower Quartile .	32.8	44.7	60.5	86.4	35.0	68.6	93.8
Median . . .	41.3	54.7	71.8	109.1	44.9	82.6	115.3
Upper Quartile	49.7	66.7	86.0	136.3	53.2	98.0	134.2

In order to facilitate comparisons the medians and interquartile ranges of the classes are set forth graphically in Chart 1.

Individual differences displayed. — This chart shows strikingly the inexorable force of individual differences and the subordinate importance of the time-factor with regard to the achievement of students, and their proper classification for instructional purposes, in modern language classes.

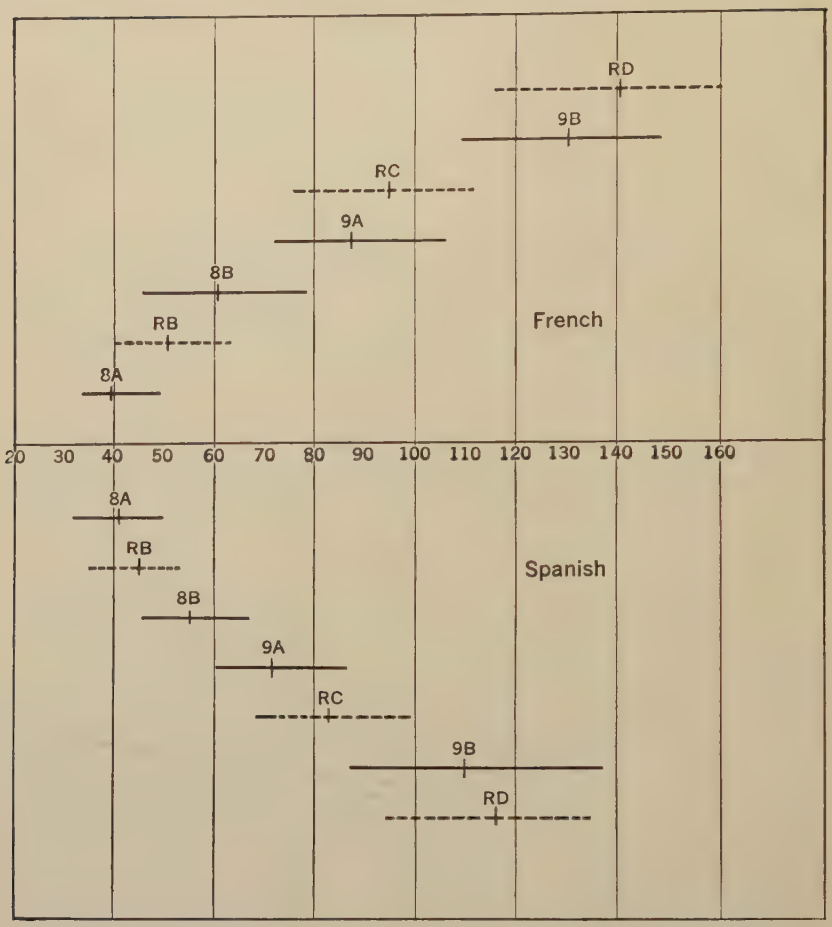


CHART 1. — Medians and interquartile ranges of scores of each class in French and Spanish in the junior high schools of New York City, June, 1925. The numbers in the center represent raw scores on both tests. The projection lines above the numbers relate to French, and those below to Spanish classes. Solid lines represent normal, and dash-lines rapid advancement classes.[1]

[1] The reader should be warned at once against making comparisons between the achievements of the French and the Spanish departments on the basis of this chart. The Spanish classes *seem* to be closer together and nearer to the lower end of the score-scale; since the two tests were constructed in the same manner and were both based mainly upon the most frequent 2000 words and upon the common fundamentals of grammar in each language the reader may be inclined to accept this *apparent* superiority of achievement in French as real. Those who make this interpretation do so at their own risk, because we do not know the relation between the units of the French and of the Spanish tests. The units of the Spanish test may have the same relation to the French test units as inches have to centimeters, or the reverse. Various considerations, too involved for brief exposition, lead the writer to infer that the units on the two tests are very nearly equal, with the chances in favor of the theory that the Spanish units are slightly larger than the French units. But we cannot know, with the data at hand, the exact relations between these two tests, and therefore the reader is invited to make comparisons between projection-lines on one side of the line of raw scores, but not on opposite sides.

(1) The medians of the classes are separated by increasing intervals as we go up the scale, the difference (in terms of raw score-units) between the 9A and 9B medians being slightly more than twice as large in both languages as the difference between the 8A and 8B medians. This seems to be due almost entirely to the progressive elimination of students because of low scholarship in general, and not because of particular deficiency in modern language work.

(2) The rapid advancement classes consistently achieve more than the normal classes and in less time. When we consider that the rapid advancement pupils were selected not on the basis of particular aptitude for or interest in modern language work, but on the basis of general scholarship, their consistent outdistancing of the normals (in only 75 per cent of the time taken by the normals) is a fact the significance of which for educational philosophy and administration can hardly be overestimated.

(3) The chart gives more than a complete justification of the provision for individual differences inherent in the rapid advancement classes, but it shows definitely the inadequacy of the method of selecting students for rapid classes, and the inadequacy of the "two-track" or "two-speed" plan of the course. This inadequacy has been recognized by Director Greenberg and by the corps of modern language teachers in the junior high schools. Every one has known that some students are able to achieve more, and do achieve more, in one semester than many others can or do achieve in four or five semesters. The need for more flexible adjustments to individual capacities and achievements has been clearly recognized, but there has been no means for securing the exact and comparable measurements of the achievements of all students in all classes at the end of each semester which are prerequisite to individual adjustments and readjustments. The teachers, in spite of the relief derived from the differentiation into normal and rapid classes, have repeatedly complained of the heterogeneity of their classes. Individual schools have done what they could to reorganize their classes so as to make possible at least a certain minimum efficiency of teaching; but a different examination or standard was used for each class and no two schools used the same examinations, except those given at the end of the course, with the result that standards in different schools and in different classes in the same school became a heterogeneous maze from which the only escape lay in a centrally administered system of standardized examinations, adapted to the whole range of modern language achievement in the junior high schools. Such a system of examinations will not solve the adjustment problem, but it will furnish that minimal information about individual achievement without which no administration, however wise, can build an educational ladder flexible enough to insure continuous adjustments to individual needs.

Student eliminations. — The number of French 9B and RD students who took the test in June, 1925, was 4315; and of 8A and RB students, 6502; that is, the number of students in the last junior high school semester class is only about 66 per cent of the number in the first. The corresponding numbers for Spanish were 2336 and 1409, which give a "survival" percentage of about 61 per cent. No analysis of eliminations will be attempted; but it seems worth while to call attention to two facts: (a) the elimination rate appears to be somewhat greater in the normal than in the rapid classes; and (b) the students eliminated appear to be a very heterogeneous group with respect to modern language achievement; that is, they are eliminated on *general* grounds, perhaps partly by chance, rather than on the basis of modern language deficiency. The "chance" nature of the eliminations is indicated by the increasing range of scores in semester classes from 8A and RB to 9B and RD classes in Chart 1; and the more extensive eliminations in normal classes are indicated in Table 2, below. The 9B French class is only 65 per cent as large as the 8A French class; while the RD French class is 70 per cent as large as the RB French class; and the corresponding figures for Spanish are 48 per cent and 85 per cent. As already indicated above, these figures should not be taken without caution; they indicate only the *gross apparent* elimination rate. The most important fact in this connection is that some good modern language students are eliminated, and many hopeless failures are retained.

TABLE 2

GROSS APPARENT ELIMINATIONS FROM MODERN LANGUAGE CLASSES IN JUNIOR HIGH SCHOOL AS INDICATED BY EXPRESSING THE NUMBERS OF STUDENTS IN 2ND, 3RD, AND 4TH SEMESTER CLASSES AS PER CENTS OF THE NUMBERS IN BEGINNING SEMESTER CLASSES.

French

	8A	8B	9A	9B	RB	RC	RD
Number of students, June, 1925 . . .	3736	3041	2604	2417	2766	2415	1898
Approximate per cent . . .	100%	81%	70%	65%	100%	89%	70%

Spanish

	8A	8B	9A	9B	RB	RC	RD
Number . . .	1486	1332	658	688	850	683	721
Per cent . . .	100%	90%	45%	48%	100%	80%	85%

French 8A + RB = 6502; 9B + RD = 4315; total French apparent survival = 66%
Spanish 8A + RB = 2336; 9B + RD = 1409; " Spanish " " = 61%

TABLE 3

French

DISTRIBUTIONS OF SCORES OF NORMAL AND RAPID ADVANCEMENT CLASSES IN FRENCH IN THE JUNIOR HIGH SCHOOLS OF NEW YORK CITY. NOTE THE OVERLAPPING: SOME STUDENTS IN 9B ARE BELOW THE AVERAGE OF 8A, AND SOME 8A STUDENTS ARE ABOVE THE AVERAGE OF 9B. EACH CLASS REPRESENTS THE WHOLE GAMUT OF ACHIEVEMENT. THE RD CLASS, IN SPITE OF ITS HIGH AVERAGE, INCLUDES SOME STUDENTS WHO ARE BELOW 8B AVERAGE, AND MANY WHO ARE BELOW 9A AVERAGE IN ACHIEVEMENT.

SCORES	8A	8B	9A	9B	RB	RC	RD	TOTAL	RECLASSIFICATION
0–2									
3–7	12				4			16	Eliminated from
8–12	88	3			33			124	modern language
13–17	139	10		1	47	1		198	classes or repeat
18–22	267	28	6		73	3		377	8A
23–27	327	58	4	2	122	7		520	
28–32	395	93	10	3	150	6		657	
33–37	485	156	27	2	194	22		886	
38–42	525	188	33	3	278	33	1	1062	8A to 8B
43–47	401	279	64	7	276	36	1	1064	
48–52	353	276	66	3	315	68	8	1089	
53–57	227	272	84	12	274	72	3	944	
58–62	157	271	112	10	258	73	8	889	
63–67	100	249	161	13	189	99	13	824	
68–72	68	235	167	22	152	119	18	781	8B to 9A
73–77	56	187	209	25	128	125	12	742	
78–82	38	143	200	42	77	140	19	659	
83–87	35	113	185	63	66	183	35	680	
88–92	18	100	205	60	50	184	42	659	
93–97	14	100	181	86	27	175	48	631	
98–102	7	69	170	91	21	189	46	593	9A to 9B
103–107	6	50	134	114	12	167	71	554	
108–112	5	49	131	112	5	153	64	519	
113–117	3	35	104	145	4	145	82	518	
118–122	2	23	95	148	4	120	87	479	
123–127		21	75	174	1	82	108	461	
128–132	1	9	55	177	2	65	107	416	
133–137	1	10	47	172		44	120	394	9B to 2nd-year
138–142		4	24	157		39	130	354	senior high school
143–147	1	2	23	161		25	123	335	
148–152	1	2	11	162	1	15	105	297	
153–157	1	1	6	124		5	138	275	
158–162		2	9	90		5	119	225	
163–167			1	88	1	7	112	209	
168–172	1		2	54		2	96	155	
173–177		1	1	45		2	66	115	
178–182			1	18			49	68	
183–187				17		2	35	54	
188–192	1		1	11			23	36	
193–197				2	1		6	9	
198–202	1			1		1		3	
203–207		2					2	4	
208–212						1	1	2	
N	3736	3041	2604	2417	2766	2415	1998	18,877	
Q_1	29.28	47.02	72.49	110.01	39.19	65.60	118.2		
Mdn.	39.48	60.91	87.30	129.99	51.25	94.04	140.15		
Q_3	50.28	77.35	108.00	148.25	64.32	111.56	159.45		

TABLE 4

Spanish

DISTRIBUTIONS OF SCORES OF NORMAL AND RAPID ADVANCEMENT CLASSES IN SPANISH IN THE JUNIOR HIGH SCHOOLS OF NEW YORK CITY. NOTE THE OVERLAPPING; SOME STUDENTS IN 9B ARE BELOW THE AVERAGE OF 8A, AND SOME 8A STUDENTS ARE ABOVE THE AVERAGE OF 9B. EACH CLASS REPRESENTS THE WHOLE GAMUT OF ACHIEVEMENT. THE RD CLASS, IN SPITE OF ITS HIGH AVERAGE, INCLUDES SOME STUDENTS WHO ARE BELOW 8B AVERAGE, AND MANY WHO ARE BELOW 9A AVERAGE IN ACHIEVEMENT.

SCORES	8A	8B	9A	9B	RB	RC	RD	TOTAL	RECLASSIFICATION
0–2	1							1	Eliminated from modern language classes or repeat 8A
3–7	8							8	
8–12	7	1			8			16	
13–17	30	3	1		39			73	
18–22	72	11	4		44			131	
23–27	124	28	1		34	1		188	
28–32	151	57	4	1	48		1	262	8A to 8B
33–37	202	73	7		94	6		382	
38–42	224	113	12		104	5	1	459	
43–47	228	135	23		145	16	2	549	
48–52	185	186	39	5	110	14	6	545	8B to 9A
53–57	87	174	40	4	92	30	5	432	
58–62	43	130	73	4	55	34	16	355	
63–67	36	123	68	14	25	58	13	337	
68–72	20	80	75	33	19	64	22	313	9A to 9B
73–77	11	60	65	27	13	58	34	268	
78–82	14	35	61	32	3	60	23	228	
83–87	5	27	49	40	8	52	22	203	
88–92	10	16	45	48	4	42	30	195	
93–97	2	10	23	35	1	68	31	170	
98–102	6	13	15	51	2	51	37	175	
103–107	4	15	11	41	1	40	48	160	9B to 2nd-year in senior high school
108–112	2	11	15	42		25	42	137	
113–117	2	4	9	54		14	59	142	
118–122	1	7	8	32		10	59	117	
123–127	1	4	3	32		15	36	91	
128–132	1	6		33		6	43	89	
133–137	1	3	3	24		4	46	81	
138–142	1	2	1	30		1	28	63	
143–147		1		30		4	30	65	
148–152	1	1	1	24		2	21	50	
153–157	1			17			23	41	
158–162			1	13		1	10	25	
163–167				2	1		11	14	
168–172				5		1	5	11	
173–177	1			2			3	6	
178–182	1	1		4			2	8	
183–187	1	1		3		1	2	8	
188–192	1	1		2			2	6	
193–197	1		1	1			2	5	
198–202							1	1	
203–207				1			2	3	
208–212				1			2	3	
213–217							1	1	
223–227				1				1	
N	1486	1332	658	688	850	683	721	6418	
Q_1	32.29	44.74	60.30	89.25	35.10	68.47	94.13		
Mdn.	41.31	54.69	71.80	109.07	44.86	82.63	115.33		
Q_3	49.82	66.58	85.09	131.18	53.61	98.39	145.17		

Distributions of scores by classes. — Tables 3 and 4 show in greater detail some of the facts brought out above, and expose the extreme heterogeneity of the classes so strikingly that comment is unnecessary. Chart 1 was presented above to justify the rapid advancement classes at the outset: Tables 3 and 4 show the need for extension of the rapid advancement idea on the basis of objective, reliable, and comparable measures of individual achievement. The inadequacy of the classification as evidenced in these tables is such that maximum efficiency of teaching is impossible.

Students reclassified on uniform basis. — While we have Tables 3 and 4 in mind we may consider in detail the method of transmuting scores on these tests into grades and the constructive principle underlying that method. The last two columns in Tables 3 and 4 give an approximate picture of the reclassification scheme which was authorized by the Director of Foreign Languages and which was adopted by all the junior high schools in so far as administrative exigencies permitted.

(1) Each student that received a score of less than 28 on the French test (Table 3), regardless of how many semesters he had "taken" French, was given a failing grade in French 8A: if the student receiving this failing grade had been in French work more than one semester, it was recommended that he be eliminated from modern language work, unless there was evidence that he had not had a fair chance and that he desired to continue the study of French; if he had studied French only one semester, he might perhaps be eliminated, on collateral evidence of his capacities and interests, or be permitted to repeat French 8A.

We use the phrase "*failing* grade" only because it is customary, and not because we believe the phrase a legitimate one. The only legitimate purpose of examinations is to describe achievement as exactly and as meaningfully as possible, and the only worthy use that can be made of such descriptions is in constructively guiding students so that they may realize their highest potentialities in school and in life. The "failures" of most students may really be the "failures" of their teachers and educational advisers. But the word is unnecessary and misleading, except where schools and examinations are still looked upon as punitive devices. We use the phrase only in reference to the students who after one or more semesters of work have secured scores below 28, and who therefore are the most likely candidates for transfer to other departments of school work. We use the word "eliminate" with the same reservations.

(2) Students receiving scores between 28 and 53 were given "passing" grades in French 8A, regardless of how many semesters they had "taken" French; they were considered eligible to enter the 8B class, by promotion or demotion, as the case might be. All students demoted were to be interviewed by their educational advisers with a view to possible transfer

of efforts to fields other than the modern languages. All students promoted more than one semester were to be considered as candidates for rapid advancement classes.

(3) Students receiving scores between 53 and 83 were given "passing" grades in 8B and were eligible to continue with 9A. Students who had reached this status in less than the normal time, i.e. in less than two semesters, were to be considered as candidates for rapid advancement classes; those who had reached this status only after three or more semesters were to be interviewed by their advisers for possible elimination from modern language work.

(4) Students receiving scores between 83 and 118 were given "passing" grades in French 9A and were eligible to continue with French 9B.

(5) Students achieving scores above 118 were given "passing" grades in French 9B and were eligible to continue with work equal in grade to that of the second year in the senior high school.

The reclassification scheme for Spanish students was identical with that for French students, and need not be specifically described. Only the score-limits in the preceding five paragraphs need be changed. This can be done by the reader by reference to the right-hand column of Table 4.

Fallacy of "time-serving" basis for classification. — Thus some students who had taken French only one or two semesters could have been given credit for French 9B by their principals if they had acted fully under the authorization given by the Director of Foreign Languages in the Junior High Schools. To those who have not escaped from the old "time-serving" theory of achievement, and who still regard school grades as punitive devices and as means for conferring vague and mystical "credits" and "honors" rather than as exact descriptions of achievement which are to be used for constructive educational guidance of students, this reclassification scheme may seem drastic beyond all reason. Some of the writer's colleagues were frankly sceptical; not, however, as to the wisdom of demoting 9B students who were found below the average of the 8A or 8B or 9A classes; but "giving" 9B credit to students who had "taken" only one semester of work and who had not "taken" 8B and 9A, seemed to them like giving much for little. But this scepticism is based on the mistaken notion that grades are "gifts" from the teacher to the student; whereas the teacher has no more right to "give" a grade in French than he has to give a grade in height or weight. To give a student of 9B achievement credit for 8A only, because he has studied French only one semester, is as justifiable as to force a precociously tall boy to sit at a school desk which is just high enough for the average student of his age. The students who achieve in one semester as much as average students do in four are precisely the students who are most likely to carry on with distinction the scholarly work for which the world is already indebted

to modern language scholars. Let us not punish these capable young students because they are not stupid enough to require four semesters for what they have achieved in one.

It will be noticed that we have suggested no separate grading scheme for the rapid advancement classes. Our failure to do this is by no means an oversight, but in entire consonance with the principle that *all* students should be measured on *one* scale, and that all measurements or grades should, in and of themselves, do no more (and no less) than *describe* in understandable terms the achievement of each student at the time of taking the examination. How the student should be advised to continue modern language work depends, among other things, on his present achievement and on the time he has found necessary to reach his present degree of proficiency. A student who has progressed more than the average student in equal or less time, should be put in an RC or RD class, according to his present achievement, or simply be promoted ahead to the normal class for which his present achievement makes him eligible. We need have no fear about "skipping" grades; these bright students have not *skipped* the work of the elided semesters; they *have learned* it because of superior effective ability, and probably in part to save themselves from boredom after they had learned the preceding semester's work but were still held in the same class to witness the slow progress of their "normal" classmates. The semester should not be regarded as the metronome for every performer in the school ensemble.

It appears that teachers generally exaggerate the limiting force of syllabi, and overestimate the importance of the artificial segmentation of courses of study into semesters or years or classes. Fortunately, a modern language is not actually "chopped up" as the syllabi are, and students do not absorb a modern language by semester-"chops." Students are always learning something that, according to the syllabus, they have no business to learn until a later semester, and are continually reinforcing something they have had in a preceding semester. Thus the good 8B students, while learning 8B material, learn more from a voluntary foretaste of 9A material than their slow classmates will learn by devoting a full semester to 9A work.

The important thing is that classes should be kept homogeneous, whether they are called "normal" or "rapid," and that each student be kept busy, constantly stimulated and mentally challenged by his class work and progressing at his maximum rate. Indeed, so long as students are reclassified every semester on the basis of one uniform, objective, reliable examination which gives valid measures of achievement for all classes and which is so constructed as to insure comparability of measures over a series of years, it does not matter much whether they are put in normal or rapid classes.

Need for comparable measurements in junior and senior high schools. It cannot be too strongly urged that modern language measurements in the junior high should be comparable with those in the senior high school. Junior high school students theoretically cover only the first year of senior high school work in modern languages, and are therefore eligible only for second year work in senior high school when they leave the junior high school. This may or may not be a correct guess. It seems quite likely that second year work in the senior high school is adequate for the *average* student who has "had" four semesters of junior high school work; but it seems certain that many students above the 9B average are really capable of taking third-year senior high school work, and that it is a dangerous waste to keep such students back. The only way in which the articulation between junior and senior high school can be improved is to extend the use of comparable measuring devices to the fourth year in senior high school, or at least to the third year, and to make the same provisions (or better) for individual differences in the senior as in the junior high school.

There seems to be no good reason why the same standard examinations, insuring comparability from year to year, and covering the whole range of modern language achievement in the first four senior high school years, should not be used in both junior and senior high schools. These tests have not been given to many senior high school students, but the indications are that students at the end of the second year in senior high school modern language work are far from being distinctly above the 9B and RD junior high school students.[1] The argument may be advanced that the students have forgotten their first-year work by the end of their second year in senior high school; but this argument is an unwarranted reflection on the senior high schools, and merits no consideration. The fact is that the students never learned their first-year materials, but got into second-year classes because of the use of unstandardized, subjective, and unreliable examinations which did not afford comparable measures. Too much emphasis cannot be put upon the necessity for comparable measures throughout the first three or four high school years of modern language work; it seems clear that most of the maladjustments in elementary language classes are due to the unfortunate and indefensible custom of giving separate and distinct examinations to the four elementary high school classes.

Equivalence of tests determined experimentally. — In further reference to the reclassification scheme proposed above (Tables 3 and 4), it should be noted that the same uniform standards for each semester class can be used as long as may be desired by the junior high schools. Form B of the junior high school examinations in French and Spanish are equiva-

[1] Cf. evidence given in the report of the Regents experiment, p. 208 this volume.

lent to Form A, and additional equivalent Forms can be constructed as often as desired. The equivalence of Forms A and B as to difficulty, variability, and representativeness of the various aspects of the languages was based upon careful word-counts and careful syntheses of grammatical elements and idioms, and upon experimental evidence secured by administering the materials in both Forms to large numbers of modern language students.[1] The equivalence of the usual old-type modern language examinations which are constructed from year to year and used in high schools and for college entrance purposes, depends upon the judgments of modern language scholars. That there are large variations from year to year and place to place in the old-type subjective examinations every one knows; the wonder is that the variations are not larger. It is certainly not surprising to find that sometimes third-year examinations derived in the old way turn out to be easier than second-year examinations similarly derived. Such errors of judgment are not derogatory to the makers of such tests; but denial of such errors, and an insistence on the use of types of examinations which necessarily involve such errors, when the feasibility and convenience and accuracy of other types have been demonstrated, are at least signs of extraordinary conservatism and intellectual cautiousness.

Distributions of scores by parts. — For the convenience of those who are now using, or may in the future use, these tests, and who may desire to make comparisons between groups of students on the basis of vocabulary scores alone, or reading or grammar scores alone, we present here without comment distributions, with medians and quartiles, of scores of about 2000 students, selected at random, on each Part of each test.

Overlapping of classes displayed graphically. — The four percentile curves below, Charts 2–5, are presented to show graphically the extent of the overlapping of classes, and for the convenience of other investigators who are using these tests and who may wish to compare the scores of students in other schools with the scores of New York City junior high school students. These graphs seem very difficult to read at first sight, but they are really as simple as the ogive curves are graceful.

Definition of a percentile curve. — The simplest and most effective description of the percentile curve and its uses has been prepared by Otis.[2] His book cannot be too highly recommended to those who wish to become acquainted with the fundamentals of statistical methods in education. Otis elsewhere defines a percentile curve as follows:

"A percentile curve is a smooth line having a horizontal length representing 100 per cent of the scores of any group of individuals and so drawn

[1] Cf. below, pp. 207–208.

[2] Otis, A. S.: *Statistical Method in Educational Measurements*, World Book Co., Yonkers, N. Y., 1925, pp. 53–67. Cf. pp. 154 ff. this volume, for confirmatory evidence on overlapping of classes in senior high schools.

TABLE 5

DISTRIBUTIONS OF SCORES ON EACH PART OF THE FRENCH AND SPANISH TESTS

Scores	Part I Frequencies		Scores	Part II Frequencies		Part III Frequencies	
	French	Spanish		French	Spanish	French	Spanish
0		1	0		9	37	120
2–4			1–2	6	20	134	132
5–7	9	1	3–4	30	54	166	286
8–10	35	10	5–6	59	53	158	213
11–13	36	28	7–8	61	86	206	172
14–16	46	37	9–10	87	133	149	156
17–19	107	69	11–12	119	145	149	119
20–22	116	93	13–14	122	131	163	98
23–25	140	106	15–16	137	160	98	84
26–28	157	142	17–18	162	155	68	60
29–31	151	142	19–20	112	146	90	51
32–34	152	158	21–22	95	103	93	42
35–37	149	151	23–24	99	115	64	25
38–40	107	159	25–26	76	102	66	26
41–43	99	149	27–28	82	87	56	23
44–46	91	138	29–30	66	67	49	27
47–49	79	134	31–32	60	72	58	32
50–52	87	103	33–34	81	68	46	33
53–55	68	82	35–36	73	62	45	22
56–58	64	68	37–38	57	53	34	14
59–61	62	63	39–40	68	63	22	22
62–64	62	35	41–42	46	38	29	12
65–67	52	36	43–44	45	31	18	10
68–70	57	37	45–46	43	16	13	5
71–73	49	29	47–48	56	11	18	2
74–76	39	14	49–50	54	11	7	2
77–79	27	4	51–52	61	5	6	3
80–82	9	4	53–54	38	3	4	1
83–85	1	2	55–56	35	3		3
86–88		5	57–58	13	2		
89–91		2	59–60	10	3		
92–94	1	4	61–62				
95–97	1		63–64				1
98–100		1					
N	2053	2007		2053	2007	2053	2007
Q_1	26.46	29.32		15.44	13.03	7.17	3.35
Mdn.	35.96	39.24		23.64	19.41	13.34	7.61
Q_3	52.27	49.71		37.41	29.24	23.90	15.21

that any point on the curve has a height representing the amount of a given score and a horizontal position on the graph representing the per cent of the scores of the group that is exceeded by the given score. A percentile curve shows at a glance not only the median score of a class but also the range and variability of the scores. It shows at a glance just what per cent of the scores of the class is exceeded by the score of any given individual and just what per cent of the class attains or exceeds any given score. Two or more curves on the same graph show very vividly the amount of overlapping of the scores of different classes."

From these charts the reader may learn, for example, what proportion of a given class, e.g., French 8B, is (1) below the average of the next lower, or (2) above the average of the next higher class in achievement; or what proportion of the 8B class is (3) nearer to the 8A average than to its own average, and what proportion is (4) nearer to a higher class average than to its own. To answer the first question, apply any straight-edge, which is kept constantly parallel to the horizontal lines in the chart, to the point in Chart 2 at which the 8A curve crosses the vertical line headed 50, that is, the 50th-percentile line; note where this straight-edge cuts the 8B curve, and read vertically above this point the per cent of 8B pupils who are below the average of 8A students — in this case about 13 per cent.

To answer the second question, slide the straight-edge up to the point where the 9A curve crosses the 50th-percentile line, being careful to keep the edge parallel to the horizontal lines in the chart; note where the straight-edge cuts the 8B curve, and read vertically above this point the proportion of 8B students who are *below* the 9A average, in this case about 83 per cent; subtract 83 from 100 to find the proportion of 8B students who are above the 9A average, i.e., 17 per cent.

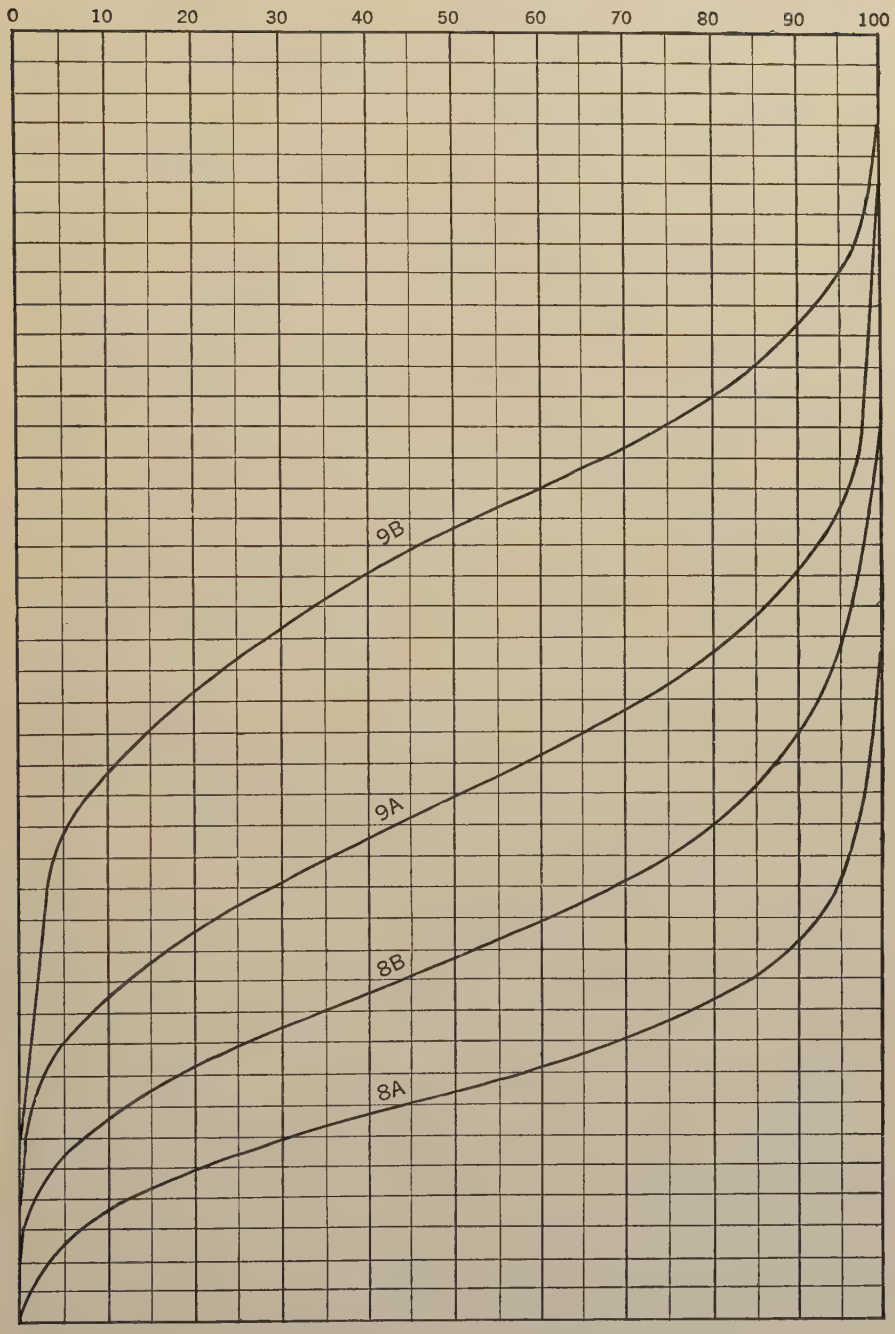
From these two simple operations we have learned the highly important fact that about 30 per cent of the 8B students in the junior high schools of New York City at the end of the 1924–1925 session were either below the average of the 8A class or above the average of the 9A class in achievement. This fact means that the 8B teachers did not have *classes* to teach, but heterogeneous aggregations of unhappy students, some of whom were below the lower quartile of the 8A class, and some of whom were above the average of the 9B class! To expect teachers or students to be at maximum efficiency in such "classes," especially when the enrolment averages about 50 students, and often reaches 75 per recitation group, is over-optimistic, to say the least.

To answer the third question suggested above, apply the straight-edge at the point midway between the points at which the 8A and 8B curves cut the 50-percentile line; keeping the edge parallel to the horizontal lines on the chart, note the point at which the edge cuts the 8B curve, and read vertically above that point the per cent of 8B students who are nearer to the 8A average in achievement than to their own class average, that is, about 30 per cent. (Holding the edge in this position and reading above the point at which it cuts the 8A curve to the right, the reader learns that about 24 per cent of the 8A students are nearer to the 8B, or some higher average, than to their own class average.)

To answer the fourth question, slide the edge up to the point midway between the points at which the 8B and 9A curves cross the 50-percentile line, taking care to keep it parallel to the horizontal lines in the graph;

Score	8A		8B		9A		9B	
	Subtotal	Percent	Subtotal	Percent	Subtotal	Percent	Subtotal	Percent
207-203			3041	100.0				
202-198	3736	100.0	3039	99.9			2417	100.0
197-193	3735	99.9	3039	99.9			2416	99.9
192-188	3735	99.9	3039	99.9	2604	100.0	2414	99.8
187-183	3734	99.9	3039	99.9	2603	99.9	2403	99.4
182-178	3734	99.9	3039	99.9	2603	99.9	2386	98.7
177-173	3734	99.9	3039	99.9	2602	99.9	2368	97.9
172-168	3734	99.9	3038	99.8	2601	99.8	2323	96.1
167-163	3733	99.9	3038	99.8	2599	99.7	2269	93.8
162-158	3733	99.9	3038	99.8	2598	99.7	2181	90.2
157-153	3733	99.9	3036	99.8	2589	99.4	2091	86.5
152-148	3732	99.9	3035	99.7	2583	99.2	1967	81.4
147-143	3731	99.8	3033	99.7	2572	98.8	1805	74.6
142-138	3730	99.8	3031	99.6	2549	97.8	1644	68.0
137-133	3730	99.8	3027	99.5	2525	96.9	1487	61.5
132-128	3729	99.8	3017	99.1	2478	95.1	1315	54.4
127-123	3728	99.8	3008	98.9	2423	93.0	1138	47.1
122-118	3728	99.8	2987	98.2	2348	90.1	964	39.9
117-113	3726	99.7	2964	97.1	2253	86.5	816	33.8
112-108	3723	99.6	2929	96.2	2149	82.5	671	27.8
107-103	3718	99.5	2880	94.7	2018	77.4	559	23.1
102-98	3712	99.4	2830	93.0	1884	72.3	445	18.4
97-93	3705	99.2	2761	90.8	1714	65.8	354	14.6
92-88	3691	98.8	2661	87.5	1533	58.8	268	11.1
87-83	3673	98.3	2561	84.2	1328	51.0	208	8.6
82-78	3638	97.4	2448	80.5	1143	43.9	145	6.0
77-73	3600	96.3	2305	75.8	943	36.2	103	4.3
72-68	3544	94.9	2118	69.6	734	28.2	78	3.2
67-63	3476	93.0	1883	61.9	567	21.8	56	2.3
62-58	3376	90.4	1634	53.7	406	15.6	43	1.8
57-53	3219	86.1	1363	44.8	294	11.3	33	1.4
52-48	2992	80.1	1091	35.9	210	8.1	21	.9
47-43	2639	70.6	815	26.8	144	5.5	18	.7
42-38	2238	59.9	536	17.6	80	3.1	11	.5
37-33	1713	45.8	348	11.4	47	1.8	8	.3
32-28	1228	32.9	192	6.3	20	.8	6	.2
27-23	833	22.3	99	3.3	10	.4	3	.1
22-18	506	13.5	41	1.3	6	.2	1	.04
17-13	239	6.4	13	.4			1	.04
12-8	100	2.7	3	.1				
7-3	12	3						

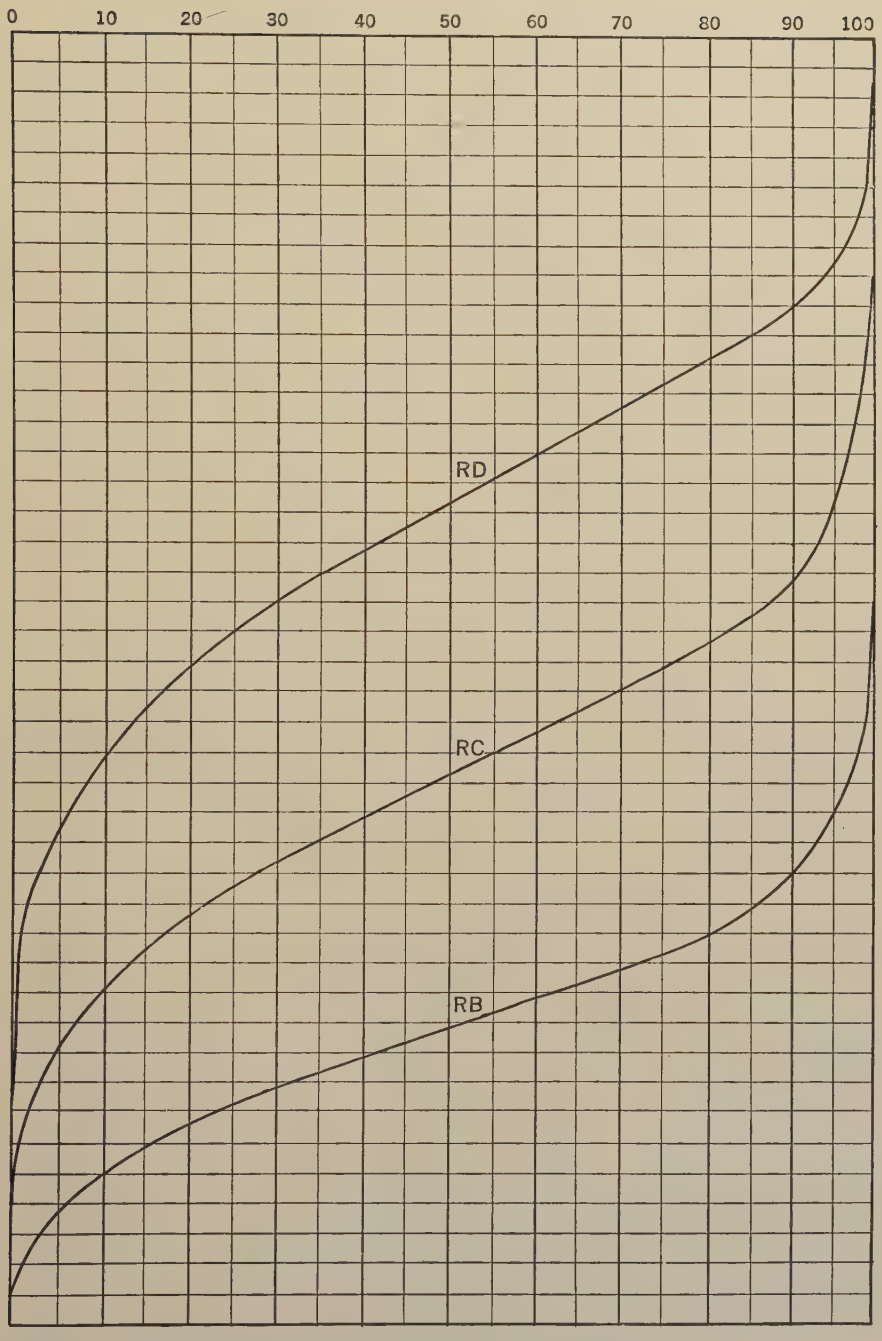
CHART 2. — Percentile curves of scores of the 8A, 8B, 9A, and 9B



French classes in the junior high schools of New York City, June, 1925.

Score	RB Subtotal	RB Percent	RC Subtotal	RC Percent	RD Subtotal	RD Percent
212-208			2415	100.0	1898	100.0
207-203			2414	99.9	1897	99.9
202-198			2414	99.9	1895	99.8
197-193	2766	100.0	2413	99.9	1895	99.8
192-188	2765	99.9	2413	99.9	1889	99.5
187-183	2765	99.9	2413	99.9	1866	98.3
182-178	2765	99.9	2411	99.8	1831	96.5
177-173	2765	99.9	2411	99.8	1782	93.9
172-168	2765	99.9	2409	99.7	1716	90.4
167-163	2765	99.9	2407	99.6	1620	85.3
162-158	2764	99.9	2400	99.3	1508	79.4
157-153	2764	99.9	2395	99.2	1389	73.2
152-148	2764	99.9	2390	98.9	1251	65.9
147-143	2763	99.9	2375	98.3	1146	60.4
142-138	2763	99.9	2350	97.2	1023	53.9
137-133	2763	99.9	2311	95.7	893	47.1
132-128	2763	99.9	2267	93.9	773	40.7
127-123	2761	99.8	2202	91.2	666	35.1
122-118	2760	99.8	2120	87.8	558	29.4
117-113	2756	99.6	2000	82.8	471	24.8
112-108	2752	99.5	1855	76.8	389	20.5
107-103	2747	99.3	1702	70.5	325	17.1
102-98	2735	98.9	1535	63.5	254	13.4
97-93	2714	98.1	1346	55.7	208	10.7
92-88	2687	97.1	1171	48.5	160	8.4
87-83	2637	95.3	987	40.9	118	6.2
82-78	2571	93.0	804	33.3	83	4.4
77-73	2494	90.1	664	27.5	64	3.4
72-68	2366	85.5	539	22.3	52	2.7
67-63	2214	80.0	420	17.4	34	1.8
62-58	2025	73.2	321	13.3	21	1.1
57-53	1767	63.9	248	10.3	13	.7
52-48	1493	54.0	176	7.3	10	.5
47-43	1178	42.6	108	4.5	2	.1
42-38	902	32.6	72	3.0	1	.05
37-33	623	22.5	39	1.6		
32-28	429	15.5	17	.7		
27-23	279	10.1	11	.5		
22-18	157	5.7	4	.2		
17-13	84	3.0	1	.04		
12-8	37	1.3				
7-3	.4	.1				

CHART 3. — Percentile curves of scores of the RB, RC, and RD French



classes in the junior high schools of New York City, June, 1925.

note where the edge cuts the 8B curve, and read vertically above that point the per cent of 8B students who are not nearer to a higher average than to their own, that is, about 71%; subtract 71 from 100 and you have the per cent of 8B students who are nearer in achievement to some higher class-average than they are to their own class-average: about 29%. (Holding the edge in this position, and reading vertically above the point where it cuts the 9A curve to the left, we find that about 30% of the 9A students are nearer to the 8B average than to their own class-average.)

From the first two questions we learned that about 30% of all 8B students are either below the average of the next lower class or above the average of the next higher class; from the answers to the last two questions, we learn that about 60% of the 8B students are nearer to some other class-average than they are to their own class-average in achievement. Thus it appears that only about 40% of the 8B students are closer to their own class-average than to some other class-average.

By reading similarly from the percentile graph for the rapid advancement classes in French, we learn that about 20% of the RC students are nearer to the RB class-average than to the RC class-average, and that about 23% are nearer to the RD class-average than to the RC class-average. At the same time we learn that 17% of the RB and about 22% of the RD students are nearer to the RC class-average than to their own respective class-averages. About 4% of the lowest rapid class is above and about 7% of the highest rapid class is below the average of the middle rapid class; about 5% of the middle rapid class is below the average of the lowest and about 6% is above the average of the highest rapid class.

From the graphs for all French classes, therefore, we learn that over 2000 of the nearly 19,000 junior high school French students are misplaced by a whole semester or more, and that approximately 10,000 of them are nearer to a higher or lower class-average than to the average of the class in which they were reciting during the spring of 1925. Teachers who have not struggled with such heterogeneity in large classes, or who have never experienced the delight of working with reasonably homogeneous classes, can scarcely apprehend the full extent of the evils, — in terms of the sacrifice of bright students on the altar of mediocrity, in terms of the cruel and more than useless browbeating of dull students, and in terms of wasted energies and frayed nerves of teachers, — which are inevitable concomitants of such misplacement and misguiding of students as this survey has uncovered in the junior high schools of New York City. Mr. Greenberg's courage in uncovering this situation, and his determination to give both teachers and students a fair chance are to be commended. The situation is only aggravated by the prevalence of different standards for each class among the various schools; but be-

fore taking up the variable standards of individual schools, we may summarize the figures for overlapping of Spanish classes.

Spanish classes less homogeneous than French classes. — Referring to the percentile graph for the four normal Spanish classes, we find markedly greater heterogeneity than in the French classes. Of the Spanish 8A students 30% are nearer to some higher class median than to their own class median; 15% are above the 8B average, and at least 5% are above the 9A average! Of the 8B students about 18% are below the 8A average, and about an equal proportion are above the 9A average! In other words, about 36% of the 8B Spanish students are misplaced by a whole semester or more, the upward misplacements being almost equal in extent and number to the downward misplacements. About 30% of the 8B students are nearer to the 8A and more than 30% are nearer to the 9A or some higher median than to their own class-average; fewer than 40% of the Spanish 8B students are nearer to their own median than to a higher or lower median!

Of the 9A students about 16% are below the 8B, and about 5% below the 8A average; about 6% are above the 9B median; and about 50% are nearer to a higher or lower median than to their own median.

Among the rapid Spanish classes, the overlapping is not so great numerically, but is probably more serious in its consequences and implications; for these classes were formed specifically to allow for individual differences, to eliminate overlapping and provide conditions favorable to the optimum development of each student. We find that about 11% of the RB students are nearer to the RC than to their own median, while about 16% of the RC students are nearer to the RB than to their own class median. About 8% of the RC students surpass the RD median; and about 25% are nearer to it than to their own class median, while about 15% of the RD students are below the RC median, and about 30% are nearer to it than to their own class median.

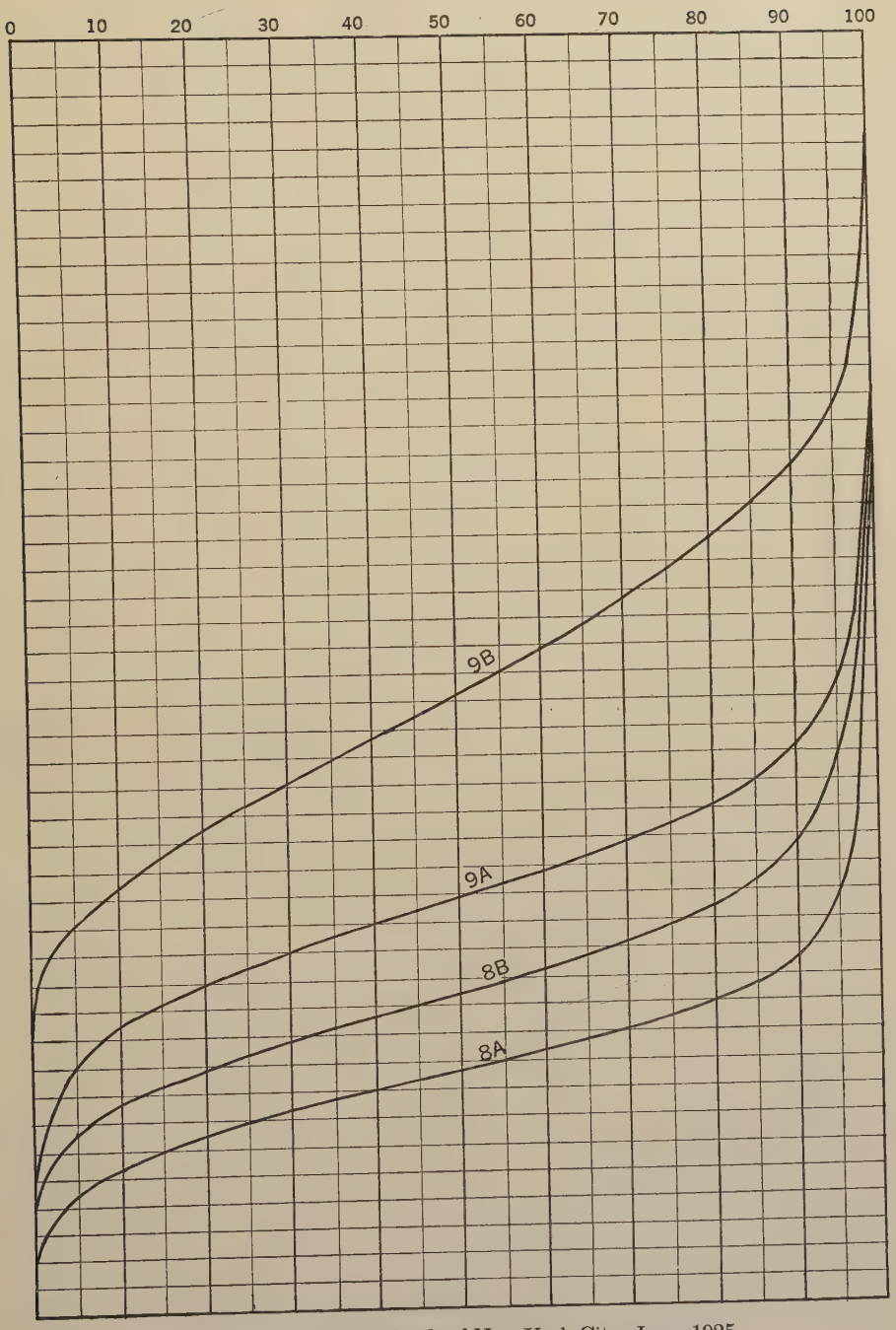
In general terms, then, more than one in ten of the Spanish students in New York City junior high schools are misplaced by one semester or more, and considerably more than half of the students are nearer in achievement to the average of a class which is one semester higher or lower than to the average of the class in which they recited during the spring of 1925.

The heterogeneity of the classes is strikingly pictured in a slightly different manner in Charts 6–9. These graphs will also be more easily understood by readers who are not familiar with percentile graphs. Charts 6–9 are simple graphs of the distributions of scores in Tables 3 and 4.

Standards in individual schools. — In Charts 10–13 the reader will find a partial explanation of the incredible overlapping of classes displayed above and will apprehend at first hand the inexorable need for *comparable*

Score	8A Subtotal	8A Percent	8B Subtotal	8B Percent	9A Subtotal	9A Percent	9B Subtotal	9B Percent
225-223							688	100.0
222-218							687	99.8
217-213							687	99.8
212-208							687	99.8
207-203							686	99.7
202-198							685	99.6
197-193	1486	100.0			658	100.0	685	99.6
192-188	1485	99.9	1332	100.0	657	99.8	684	99.4
187-183	1484	99.8	1331	99.9	657	99.8	682	99.2
182-178	1483	99.7	1330	99.8	657	99.8	679	98.7
177-173	1482	99.7	1329	99.7	657	99.8	675	98.1
172-168	1481	99.6	1329	99.7	657	99.8	673	97.8
167-163	1481	99.6	1329	99.7	657	99.8	668	97.1
162-158	1481	99.6	1329	99.7	657	99.8	666	96.8
157-153	1481	99.6	1329	99.7	656	99.7	653	94.9
152-148	1480	99.5	1329	99.7	656	99.7	636	92.4
147-143	1479	99.5	1328	99.6	655	99.5	612	89.0
142-138	1479	99.5	1327	99.6	655	99.5	582	84.6
137-133	1478	99.4	1325	99.5	654	99.4	552	80.2
132-128	1477	99.3	1322	99.2	651	98.9	528	76.7
127-123	1476	99.3	1316	98.8	651	98.9	495	71.9
122-118	1475	99.2	1312	98.5	648	98.5	463	67.3
117-113	1474	99.1	1305	98.0	640	97.2	431	62.6
112-108	1472	99.0	1301	97.6	631	95.9	377	54.8
107-103	1470	98.9	1290	96.8	616	93.6	335	48.7
102-98	1466	98.6	1275	95.7	605	92.0	294	42.7
97-93	1460	98.2	1262	94.7	590	89.7	243	35.3
92-88	1458	98.1	1252	93.9	567	86.1	208	30.5
87-83	1448	97.4	1236	92.8	522	79.3	160	23.3
82-78	1443	97.0	1209	90.7	473	71.8	120	17.4
77-73	1429	96.1	1174	88.1	412	62.6	88	12.8
72-68	1418	95.4	1114	83.6	347	52.7	61	8.9
67-63	1398	94.0	1034	77.6	272	41.3	28	4.1
62-58	1362	91.6	911	68.4	204	31.0	14	2.0
57-53	1319	88.7	781	58.6	131	19.9	10	1.5
52-48	1232	82.9	607	45.6	91	13.8	6	.9
47-43	1047	70.4	421	31.6	52	7.9	1	.1
42-38	819	55.1	286	21.5	29	4.4	1	.1
37-33	595	40.0	173	13.0	17	2.6	1	.1
32-28	393	26.5	100	7.5	10	1.5	1	.1
27-23	242	16.3	43	3.2	6	.9		
22-18	118	7.9	15	1.1	5	.8		
17-13	46	3.1	4	.3	1	.2		
12-8	16	1.1	1	.1				
7-3	9	.6						
2-0	1	.1						

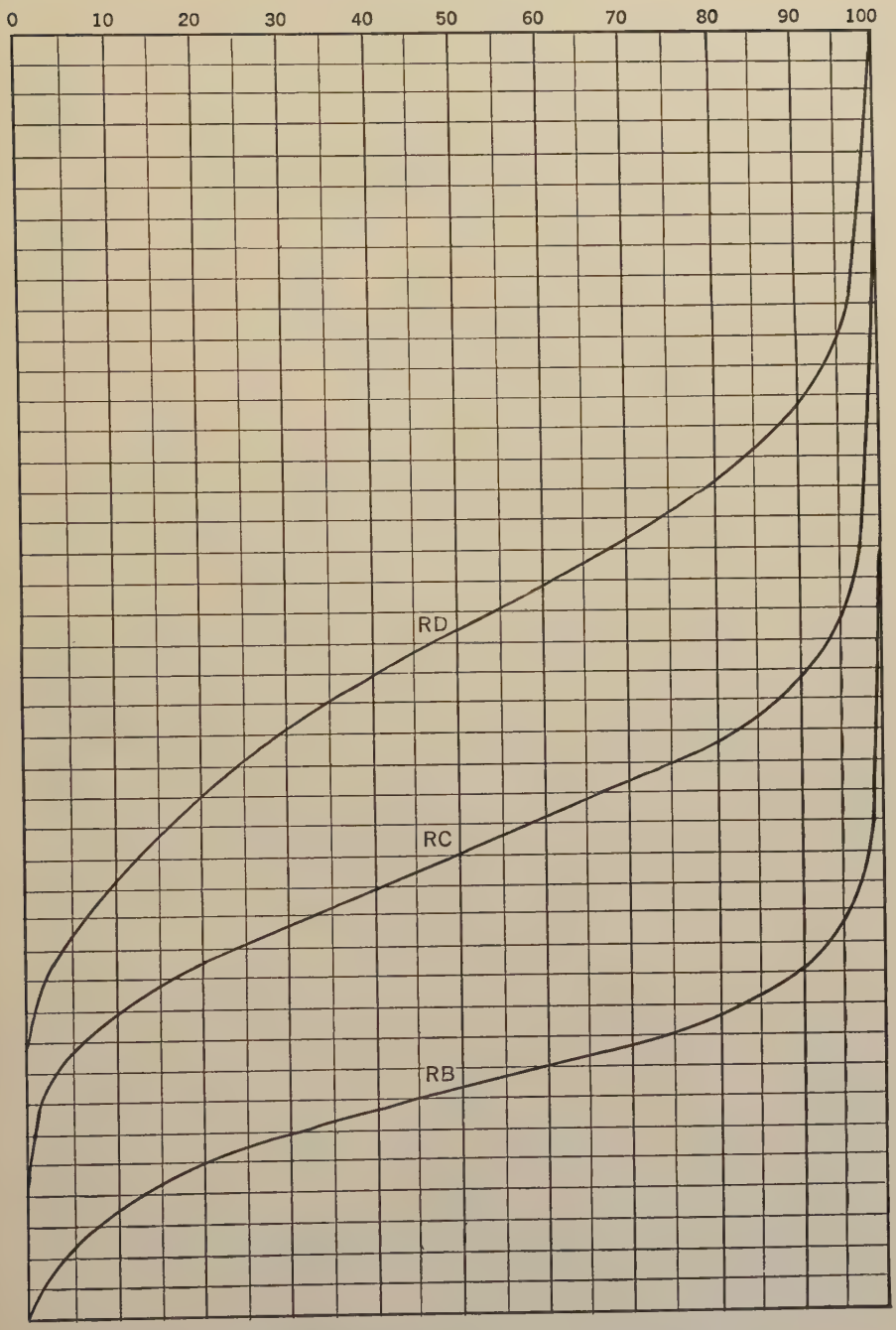
CHART 4. — Percentile curves of scores of the 8A, 8B, 9A, and 9B



Spanish classes in the junior high schools of New York City, June, 1925.

Score	RB		RC		RD	
	Subtotal	Percent	Subtotal	Percent	Subtotal	Percent
217-213					721	100.0
212-208					720	99.9
207-203					718	99.6
202-198					716	99.3
197-193					715	99.2
192-188					713	98.9
187-183			683	100.0	711	98.6
182-178			682	99.9	709	98.4
177-173			682	99.9	707	98.0
172-168			682	99.9	704	97.6
167-163	850	100.0	681	99.7	699	97.0
162-158			681	99.7	688	95.4
157-153			680	99.5	678	94.0
152-148			680	99.5	655	90.9
147-143			678	99.2	634	87.9
142-138			674	98.7	604	83.8
137-133			673	98.5	576	79.9
132-128			669	97.9	530	73.5
127-123			663	97.1	487	67.6
122-118			648	94.9	451	62.6
117-113			638	93.4	392	54.4
112-108			624	91.4	333	46.2
107-103	849	99.9	599	87.7	291	40.4
102-98	848	99.8	559	81.8	243	33.7
97-93	846	99.6	508	74.4	206	28.6
92-88	845	99.4	440	64.4	175	24.3
87-83	841	99.0	398	58.2	145	20.1
82-78	833	98.0	346	50.6	123	17.1
77-73	830	97.7	286	41.9	100	13.9
72-68	817	96.1	228	33.4	66	9.2
67-63	798	93.9	164	27.0	44	6.1
62-58	773	90.9	106	15.5	31	4.3
57-53	718	84.5	72	10.5	15	2.1
52-48	626	73.7	42	6.1	10	1.4
47-43	516	60.7	28	4.1	4	.6
42-38	371	43.7	12	1.8	2	.3
37-33	267	31.4	7	1.0	1	.1
32-28	173	20.4	1	.1	1	.1
27-23	125	14.7	1	.1		
22-18	91	10.7				
17-13	47	5.5				
12-8	8	.9				

CHART 5. — Percentile curves of scores of the RB, RC, and RD



Spanish classes in the junior high schools of New York City, June, 1925.

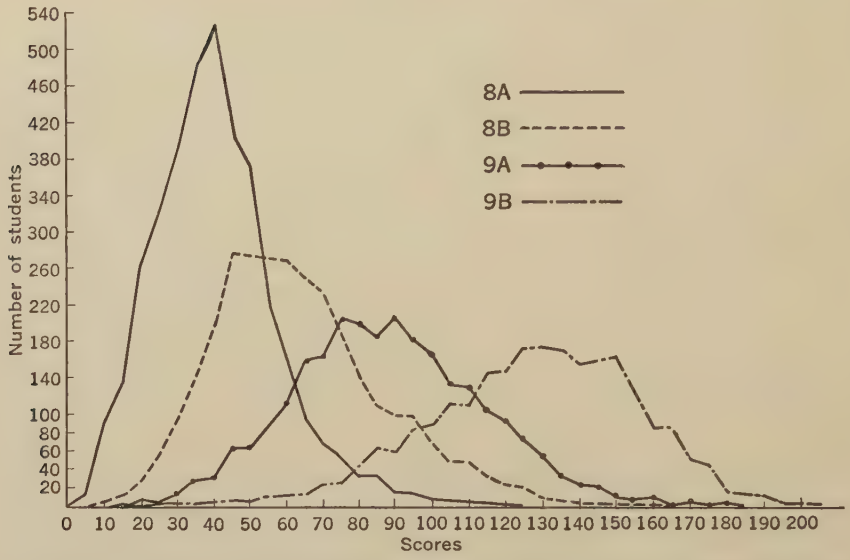


CHART 6. — Graphs of the distributions of scores of 8A, 8B, 9A, and 9B students of French in the junior high schools of New York City, June, 1925.

measures, such as are afforded by an objective test including a valid sampling of modern language elements carefully graded in difficulty to cover the whole range of achievement in the first two or three years, and which can be indefinitely duplicated by the construction of additional *equivalent* Forms.

The charts are very easy to read. The four horizontal lines in Chart 10 represent the averages of the classes on a city-wide basis, the score scale being at the left. Thus the reader will notice that the lowest line, marked at its left "8A," is drawn just a little below the line corresponding to a

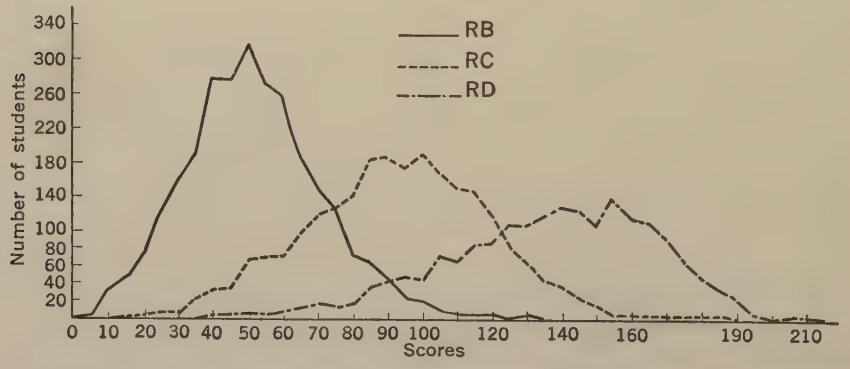


CHART 7. — Graphs of the distributions of scores of RB, RC, and RD students of French in the junior high schools of New York City, June, 1925.

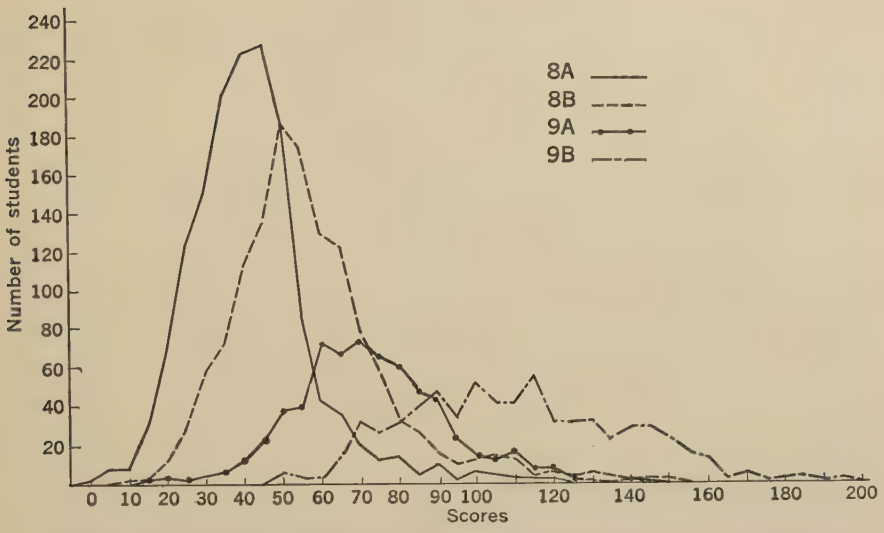


CHART 8. — Graphs of the distributions of scores of 8A, 8B, 9A, and 9B students of Spanish in the junior high schools of New York City, June, 1925.

score of 40; by referring to Table 1 the reader will find that the average of all 8A students was 39.5. The average score of all 8B students is given in Table 1 as 61, and the 8B line in Chart 10 is drawn slightly above the line corresponding to a score of 60. And so with the 9A and 9B lines.

The vertical line near the left end of each of the class average lines represents the interquartile range of the class, that is, the range of the middle fifty per cent of the students in the class. Thus the vertical line near the left end of the 9B average line extends from the line corresponding to a score of 110 to a point above the 9B average line corre-

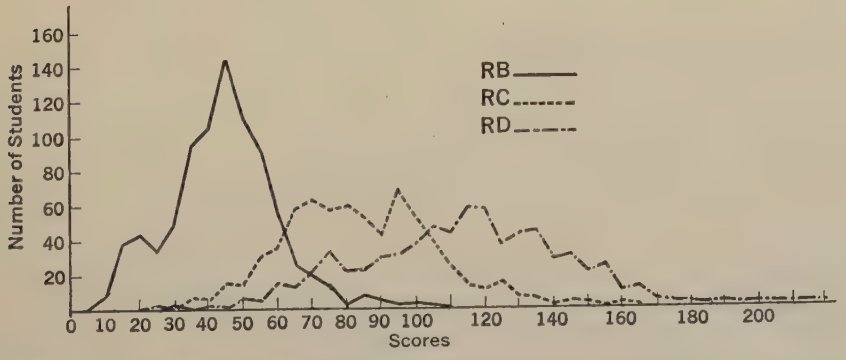


CHART 9. — Graphs of the distributions of scores of RB, RC, and RD students of Spanish in the junior high schools of New York City, June, 1925.

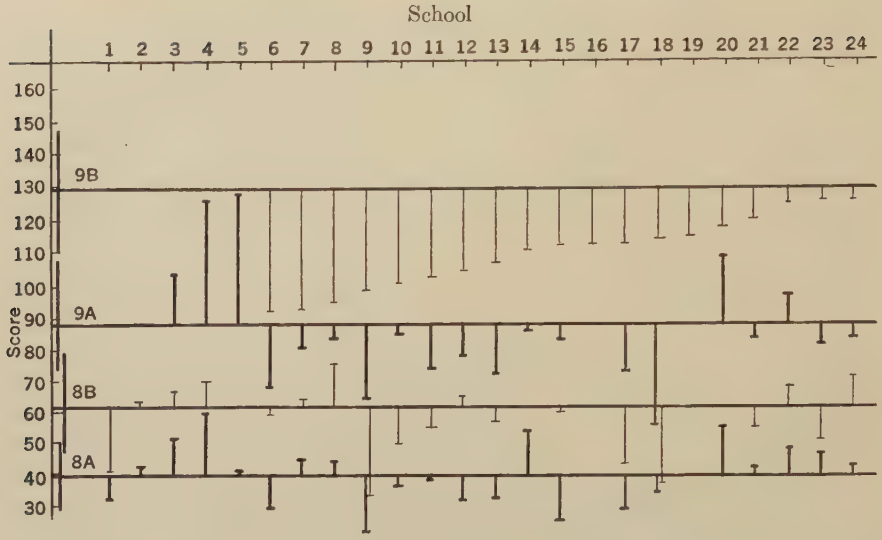
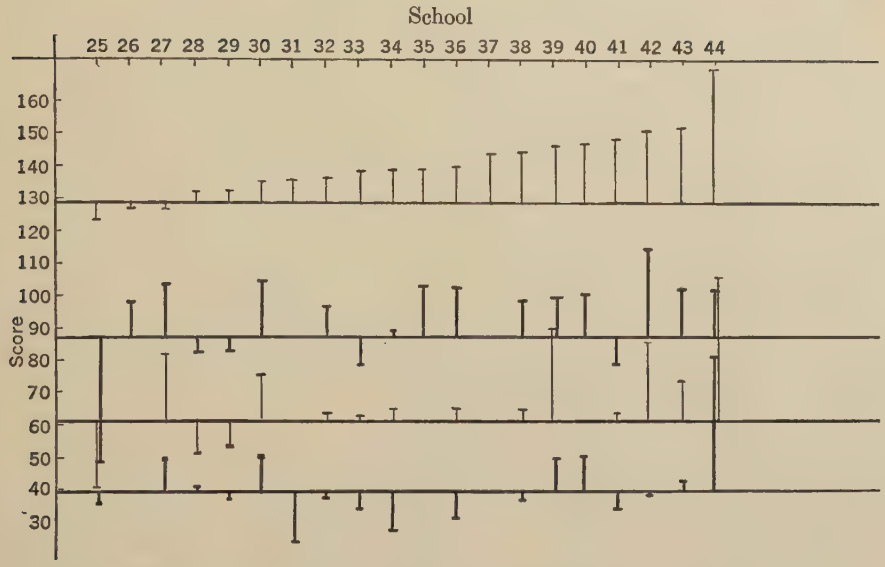


CHART 10. — Graphic comparison of median scores

sponding to a score of 148. These points are also based upon Table 1; if the reader will look at Table 1 he will find that the lower quartile of the 9B class is 110, and the upper quartile is 148. In other words, 25% of the 9B students secured scores below 110, 50% secured scores between 110 and 148, and 25% secured scores above 148. The interquartile ranges of the classes as indicated on these charts will be a convenience to certain types of readers, but they are not at all essential to an understanding of the main features of Charts 10 to 13, inclusive. Teachers who have no predilection for statistical refinements may forget that these lines are in the charts. The important thing is to keep in mind that the horizontal lines, in order, represent the class averages in terms of the original units of the tests, which may be read on the score-scale at the extreme left of each chart.

Without further explanation, let us compare the two schools whose identities in Chart 10 we have hidden under the numbers 5 and 6. Number 5 had no 8A, 8B or 9B students; it had only students which it classified as 9A. But the heavy line rising from the 9A average-line shows that the average score of this 9A class is actually 127 — almost equal to the 9B city-wide average! In school Number 6 the light line dropping down from the 9B average-line shows that the students classified as 9B secured an average score of only 91 — almost average for 9A classes! In other words, No. 5 actually had a 9B class, but called it a 9A class; whereas No. 6 had no 9B class, but called what was actually a 9A class a 9B class. Nor did school No. 6 have any 9A class other than the group of students



of each "normal" French class in each school.

it *called* 9B; for the students which this school classified as 9A are actually only slightly above the 8B city-wide average. Here we have two classes *called* 9A, one of which (School No. 5) is actually an average 9B class, and the other (No. 6) is only slightly above average for 8B classes, — a whole year apart in actual achievement.

As a further example of how the charts should be read, let us compare schools 8, 9, and 10. Notice that in these charts we have arranged the schools in an ascending order of 9B class averages, from left to right. Thus schools 8, 9, and 10 have 9B classes with approximately equal averages, all of which are nearer to the 9A city-wide average than to the 9B city-wide average. But the three schools are alike in no other respect, in so far as class standards are concerned. School 9 classes are consistently about one semester below city-wide standards; indeed, the 8B class in this school is somewhat below the city-wide 8A average. In schools 8 and 10 the 8A and 9A classes are almost exactly at par; but the so-called 8B classes in these schools vary from the city-wide average about equally in opposite directions.

In schools 18 and 20 the 8A and 9A classes vary in opposite directions. In 18 the 9A class average is below the 8B city-wide standard, while in school 20 the 9A class average is about halfway up to the 9B standard; and in 18 the 8A together with the 8B class is below par, while in 20 the 8A is almost average for 8B classes.

In Chart 11 the variation of class standards within a given school is not so great nor so frequent as in Chart 10; that is, if a given school is

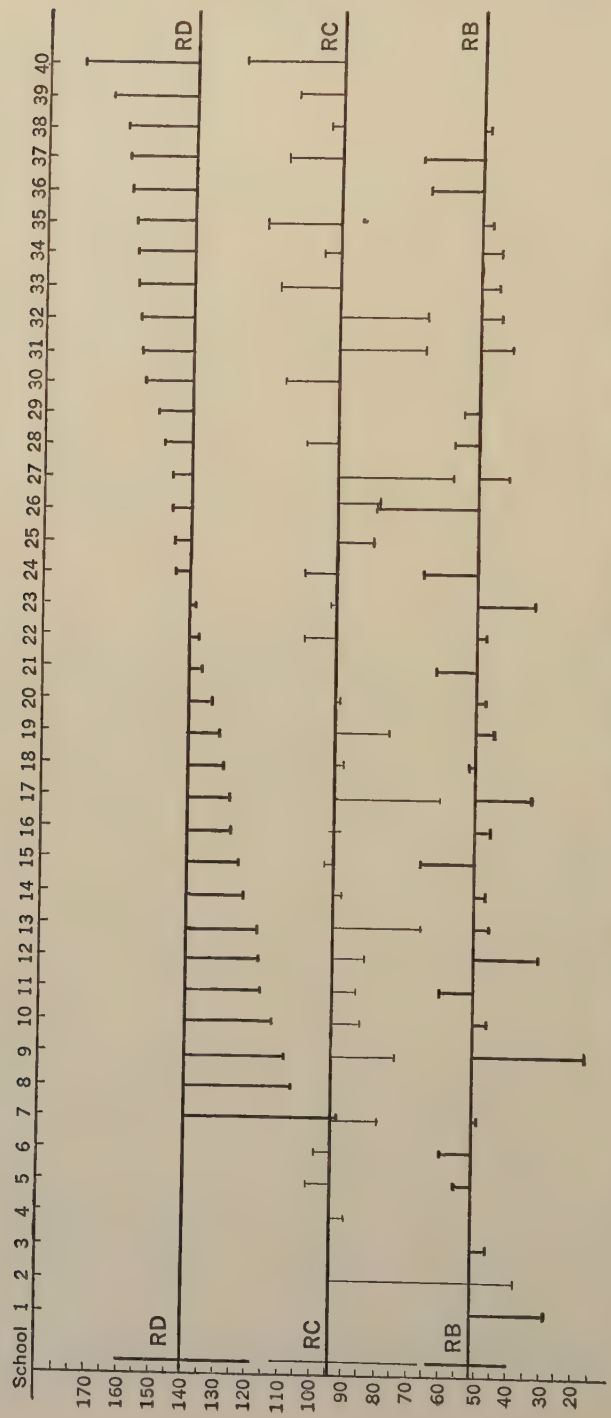


Chart 11. — Graphic comparison of median scores of each "rapid" French class in each school.

below city-wide standards for the RD class it is more likely to be below standard for RC and RB classes than above in one or both these classes; but we find the same tremendous variations between different schools. School No. 2 in Chart 11 had an aggregation of students which it classified as RC, when actually they are considerably below the RB average, — indeed, they are slightly below the 8A city-wide average, although they have had two full semesters of work and were classified as rapid advancement pupils! School 7 had an RD class which at the end of

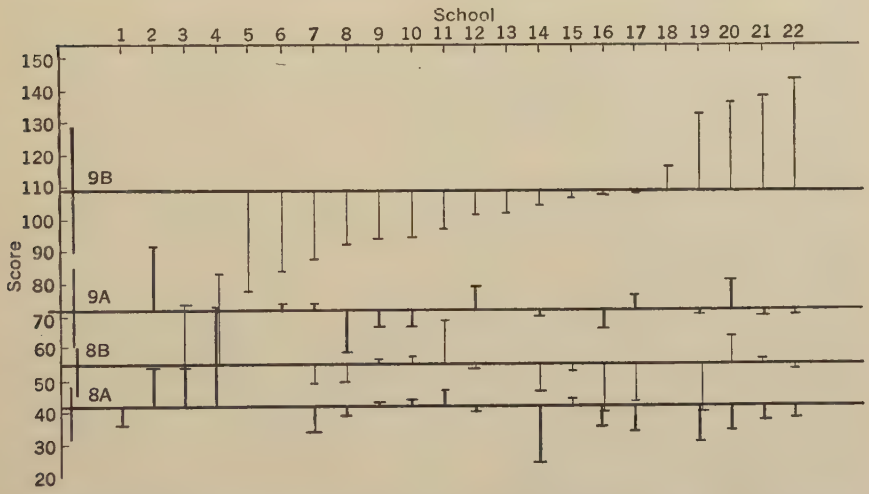


Chart 12. — Graphic comparison of median scores of each "normal" Spanish class in each school.

three semesters of "rapid" advancement is just at par for RC pupils. School 9 has an RB class which is practically entirely below the lower quartile of 8A students! Probably 90% of the students in this RB class and in the RC class in school 2 should never have been allowed to study a modern language; yet they were not only allowed to try it, but were classified as rapid advancement pupils.

Charts 12 and 13 for the normal and rapid Spanish classes tell a similar story. Some readers may be inclined to discount the variations displayed, and to attribute the *appearance* of extreme variations to imperfections in the measuring device used, and to other like causes; they will be impelled to seek these explanations by the feeling that such gross variations are really impossible. We shall now present such evidence as we have concerning the reliability and validity of the tests used. In judging the force of our statistical evidence the reader must remember that the tests were constructed with the intimate coöperation of the Director of Foreign Languages in the Junior High Schools and of some of his teachers,

and with the counsel and advice of a large number of his teachers. The tests are the product of four years of continuous experimenting with many kinds of tests. Both the content and form of each question in each test were approved by teachers who were actually teaching pupils who were to be tested. It is also to be noted that we have had no adverse criticism on the fairness of the tests from any teachers whose pupils

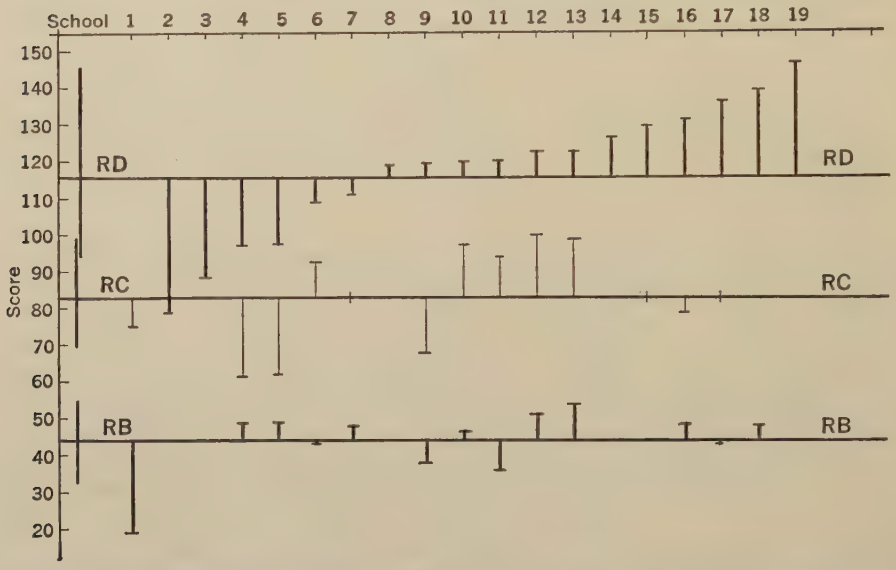


Chart 13. — Graphic comparison of median scores of each "rapid" Spanish class in each school.

were subjected to and rated by them. On the contrary, we have had many expressions of satisfaction.

As to the possibility of such extreme variations, let us remember that no uniform examinations of any sort have ever been given below the 9B and RD classes. In the absence of some common point of reference and of some common unit of reckoning it would be nothing short of a miracle if such variations did not exist. No one would be surprised if we found such variations in the subjective judgments of teachers, or even of trained physicists, as to the temperatures of a series of days; but the estimation of temperature is a simple task as compared with the matter of judging a very complex phenomenon such as modern language achievement.

III

RELIABILITY AND VALIDITY OF TESTS

Reliability of Tests

Definition of reliability. — By reliability of a test is meant the degree to which it is consistent with itself in measuring the same thing two or more times. The most direct way to find the reliability or consistency of a test is to administer two equivalent Forms of the test to the same students, and then calculate the correlation between the two sets of scores. But this procedure is not always feasible. When two equivalent Forms cannot be administered to the same students, the reliability of the test may be estimated from the reliability of one-half the test, that is, from the correlation between the two sets of scores obtained by "splitting" the one Form of the test that has been given to students. "Splitting" a test is accomplished by treating the odd-numbered questions as though they constituted a test separate and independent of the test constituted by the even-numbered items. Thus each student receives two scores on the Form that has been administered: one score is based entirely upon the odd-numbered items, and the other is based entirely upon the even-numbered questions in the examination. The correlation between these two sets of scores shows the consistency or reliability of one-half of the test; by using the Spearman-Brown Formula the reliability of the whole test may be very closely estimated from the reliability of one-half of it. This is what we have done with the junior high school French and Spanish tests.

It is easy to see that if we had an ideal test a student who received a total score of 60 would receive a score of 30 on each random half of the test; a student who received a total score of 90 would receive a score of 45 on each random half; and so with other total scores. But ideal tests do not exist, and our best course is to find out what type of test approaches most nearly to the ideal and to use it in preference to all others, remembering, of course, that reliability alone does not make a test ideal. A test must not only measure consistently and accurately, but it must measure the right thing.

Reliability of old-type examinations. — In considering modern language tests it is not our privilege to choose between perfect and imperfect instruments: we can only try to avoid the less reliable and seek the more valid and more reliable forms of examinations. The old-type of modern lan-

guage examination is of such a nature that it is very difficult to ascertain its reliability, a fact which in itself argues against its having high reliability. What evidence there is tends to show that the ordinary old-type examination of two or three hours length has an average reliability of about 0.70; it is almost certain that it is not above 0.80, and reaches 0.80 only in rare cases, if at all. These figures may be compared with the reliability coefficients found for the new-type junior high school tests.

Reliability of new-type tests. — The correlations between random halves of the new-type tests for four 500-pupil samplings range for the French from 0.905 to 0.961, and for the Spanish test from 0.882 to 0.953, with averages at about 0.95 and 0.92, respectively. The four samplings were chosen so as to represent all types of schools: the first five hundred papers were all taken from the schools with the lowest average score; the second sample was made up of examination returns from low-middle schools, the third from high-middle schools, and the fourth from the schools which averaged distinctly above the city-wide average.

Treating the two thousand cases in each language as a single sampling we find that the correlation between random halves is nearly 0.94 for the French test, and a little over 0.93 for the Spanish test. From these figures we estimate by means of the Spearman-Brown Formula that the reliability of each total test is almost exactly 0.97. For a 90-minute examination this coefficient is very gratifying. It is only 0.17 above the estimated maximum reliability for old-type modern language examinations, but the error of estimate, expressed in terms of standard deviation, with a reliability of 0.97 is only about one-third the error of estimate with a reliability of 0.80.

The correlations between random halves of each of the Parts of each test, and the Spearman-Brown estimates of total reliabilities of each of the Parts, and of the whole, of each of the two tests, are shown in Table 6. These coefficients may be accepted with confidence in view of the fact that each is based upon a random sampling of 2000 test returns.

TABLE 6

Reliabilities of French and Spanish junior high school tests. The Number of cases for each correlation is 2,000.

	French		Spanish	
	$r_{\frac{1}{2}\frac{1}{2}}$	S–B	$r_{\frac{1}{2}\frac{1}{2}}$	S–B
Part I	0.877	0.936	0.871	0.93
Part II	.911	.955	.853	.92
Part III	.927	.963	.880	.94
Total	.937	.970	.934	.97

Desirability of 500-question tests. — The close correspondence between the reliability findings for the two tests is an interesting fact in itself, and carries certain implications which seem worthy of note. In the first place, it argues that the total reliability coefficients of both tests have been determined with high accuracy. In the second place, in view of the absolute parallelism in the derivations of the two tests, it indicates both that these forms of questions can be depended upon when similarly used to give very good results in other languages, and that with these forms of questions, used as here used, we have reached the upper limit of reliability when the total number of separate items is only 220 or 225. The importance of this last indication is that it compels us to make and use longer tests than we have been accustomed to thinking of as generous in length and in sampling of materials and of performances. These indications have been more than confirmed in similar experiments in other subject matters. In modern language measurements we must become accustomed to tests of 350 to 500 separate items of the sorts used in these tests, and in about equal proportions. Tests of 350 to 500 items can easily be administered in 120 to 150 minutes, and the gain in reliability and significance will more than repay the few extra minutes used.

Reliability of each Part of the new-type tests. — The correlations in Table 7 were calculated so as to secure as many bits of independent evi-

TABLE 7

RELIABILITY COEFFICIENTS OF THE PARTS AND OF THE WHOLES OF THE JUNIOR HIGH SCHOOL TESTS, BASED ON FOUR 500-PUPIL SAMPLINGS. EACH OF THESE FOUR SAMPLINGS CONTAINS EQUAL NUMBERS OF 8A, 8B, 9A, AND 9B STUDENTS. THE FOUR SAMPLINGS REPRESENT, RESPECTIVELY, SCHOOLS WHICH TEST FAR BELOW, A LITTLE BELOW, A LITTLE ABOVE, AND FAR ABOVE THE CITY-WIDE AVERAGE FOR ALL CLASSES.

Types of Schools	Low		Middle Low		Middle High		High		Averages	
Cases	500		500		500		500		4×500	
	$r_{\frac{1}{2}\frac{1}{2}}$	S-B r_{11}	$r_{\frac{1}{2}\frac{1}{2}}$	S-B r_{11}	$r_{\frac{1}{2}\frac{1}{2}}$	S-B r_{11}	$r_{\frac{1}{2}\frac{1}{2}}$	S-B r_{11}	$r_{\frac{1}{2}\frac{1}{2}}$	S-B r_{11}
French Test										
Part I	.852	.919	.841	.913	.895	.945	.847	.917	.859	.924
Part II	.908	.952	.864	.927	.944	.972	.885	.939	.925	.962
Part III	.875	.933	.869	.930	.902	.948	.869	.930	.879	.936
I+II+III	.952	.975	.932	.965	.943	.971	.961	.980	.945	.972
Spanish Test										
Part I	.822	.902	.831	.908	.830	.908	.843	.915	.832	.908
Part II	.791	.886	.918	.957	.860	.925	.894	.944	.866	.928
Part III	.819	.901	.916	.956	.918	.957	.929	.964	.895	.945
I+II+III	.883	.938	.902	.949	.912	.954	.953	.966	.913	.955

dence on the reliability of the tests as possible. The variations between the corresponding correlations from the four samplings are, on the whole, slight, and strongly confirm the general results reported above in Table 6. However, we shall consider some of the differences in detail, partly to show how closely the correlations of Table 7 confirm those of Table 6, and partly to take the opportunity which these apparent differences offer of showing the reader concretely that correlation coefficients cannot always be accepted at face value.

Standard and probable errors of estimate. — The magnitude of a given coefficient of correlation depends, among other things, upon the range of talent, in other words, upon the spread or variability of the scores concerned in its calculation. The real agreement between two sets of scores for each of two classes may be exactly equal; but if the whole range of the talent in question is represented in one class, and only a part in the other, the correlation coefficient based on the first class will be larger than that based on data from the second class. In order to learn the true meaning of two or more correlations based upon scores of a given test, we must know, among other things, the spread or variabilities of the groups involved in the various correlations which are being compared.

A convenient way of comparing two or more such correlations is to express them all in terms of the same unit, that is, in terms of the units of the test. This is done by using the well-known formula, $\sigma_{1.2} = \sigma_1 \sqrt{1 - r_{12}^2}$ on each of the correlations which are being compared, with their corresponding appropriate sigmas. Thus, in Table 7, we find that the reliability of the whole French test is 0.952 according to one sampling of 500 students (low-average schools) and is 0.961 according to the data from 500 students in high-average schools. Substituting the appropriate values in the formula, we find that $\sigma_{1.2}$, or the standard error of estimate, according to the first sampling is 5.91, and according to the second it is 6.66, which leaves us with a very slight and negligible difference. The corresponding correlations for the Spanish test, 0.883 and 0.953, similarly reduce to standard errors of estimate of 6.33 and 5.75 raw score-units.

By this method, correlations based upon data from the same or equivalent tests, in which the scale units are equal or very nearly equal throughout the whole range of the scores involved, can be made comparable because they can be expressed in identical units. But correlations based upon data from different tests cannot be compared unless the relations of the units of the compared tests are known, or unless the compared tests have been administered to the same, or to an equally variable, group of examinees. Some writers have erroneously supposed that the reliability coefficients of *any* two tests could be compared if both were expressed as standard errors of estimate. The essence of comparability is that the quantities compared be expressed in identical units: correlation

coefficients are not directly comparable unless they are based upon data from the same or equally variable groups of examinees; standard (or probable) errors of estimate are not directly or immediately comparable unless they are expressed in terms of the same or equal scale units.

Validity of Tests

Since no other modern language test was given to the students involved in this survey, it is not possible to check the validity of the whole test against an independent criterion of achievement in modern language work. There is, however, little need for verifying the validities of Parts I and III; vocabulary and grammar tests of this sort have been so thoroughly tried out and proved that no scientist or teacher who has kept pace with recent developments can doubt their qualities. This form of vocabulary test began to be used about five years ago, and is now extensively used by modern language teachers all over the country. The grammar-completion form has been in successful use much longer; it has been considered an indispensable part of modern language examination technique by many teachers for many years. The particular form of grammar-completion test used in this survey retains all the values of older forms; and adds some values and conveniences which older types of the completion method did not possess.

Intercorrelations of Parts of new-type tests. — The only real innovation in these tests is the form of the reading-comprehension Part. As an indication of its validity we have calculated the correlations between Part II scores on the one hand, and Part I and Part III scores on the other; and as a sort of "control," we have also found the relationship between Part I and III scores. Part II of the French test gives correlations of 0.88 and 0.80 with Parts I and III, respectively; and Part II of the Spanish test gives corresponding correlations of 0.84 and 0.74. The "control" correlations between Parts I and III are 0.80 and 0.728 for the French and Spanish tests, respectively. The number of cases on which each of these coefficients is based is 2000. If we accept the validity of Parts I and III, we must accept that of Part II. On the whole, the relationships are high enough to vindicate Part II for general measurement purposes, and low enough to show that it is not a mere duplication of Parts I and III.

Table 8 shows intercorrelations between the Parts of the French and Spanish tests based upon four five-hundred-pupil samplings as well as upon the aggregate of 2000. The differences between the results from the four samplings are negligible; they are due to slight differences in the variabilities of the four groups and to the fact that some students in some schools were not compelled or encouraged to distribute their time equably between the three Parts.

TABLE 8

INTERCORRELATIONS OF THE PARTS OF THE JUNIOR HIGH SCHOOL TESTS BASED UPON FOUR 500-PUPIL SAMPLINGS AND UPON THE AGGREGATE OF THESE FOUR SAMPLINGS. EACH OF THE FOUR SAMPLINGS CONSISTS OF EQUAL NUMBERS OF 8A, 8B, 9A, AND 9B STUDENTS. THE FOUR SAMPLES WERE MADE UP SO AS TO REPRESENT RESPECTIVELY SCHOOLS WHICH TESTED FAR BELOW, A LITTLE BELOW, A LITTLE ABOVE, AND FAR ABOVE THE CITY-WIDE AVERAGE, IN ALL FOUR CLASSES.

French Test

Types of Schools	Low		Middle-Low		Middle-High		High		Unselected	
N	500		500		500		500		2000	
Parts Correlated	2	3	2	3	2	3	2	3	2	3
1	.870	.813	.876	.765	.894	.823	.876	.883	.883	.810
2		.803		.796		.854		.812		.801

Spanish test

Parts Correlated	2	3	2	3	2	3	2	3	2	3
1	.78	.623	.842	.723	.85	.728	.828	.830	.84	.737
2		.61		.774		.702		.800		.728

IV

ANALYSIS OF INDIVIDUAL QUESTIONS IN NEW-TYPE TESTS

Difficulties of Individual Questions

As was indicated on an earlier page, the materials for these tests were selected carefully so as to put preponderant emphasis on the fundamentals of vocabulary, grammar and idiom which are most nearly common to all courses of study, and at the same time provide an adequate range of difficulty, by including some questions easy enough to differentiate among the most backward students at the end of the first semester, and some difficult enough to differentiate among the best students at the end of the fourth semester. That the tests are adequate for the range of talent represented by junior high school students has been shown by the distributions of total scores in Tables 3 and 4.

Individual questions experimentally validated. — This adequacy was not achieved by the original selection alone. Large numbers of questions of each type were constructed on the basis of word-counts of widely-used textbooks, word lists in several syllabi, and syntheses of the idioms and grammatical elements in several first-year texts; then these questions were administered to an aggregate of several thousand students of French and Spanish in New York City junior and senior high schools. The absolute and relative difficulty and goodness of each individual question were determined upon the basis of experimental results. Many of the questions were eliminated for a variety of reasons; the remaining questions of each kind were then arranged in order of difficulty and divided into two equivalent Forms, A and B. Form A of the French and of the Spanish tests are reproduced below (page 48) approximately in the form in which they were administered to more than 25,000 junior high school students in June, 1925.

Table 9 shows how well we succeeded, by the end of the preliminary experimentation, in securing adequate gradations of difficulty in the items of the two tests. Other things being equal, that test is best in which the items range from very easy to very difficult and in which the increments of difficulty are *small*, *regular*, and *numerous*, thus providing a scale without gaps, like a well-built ladder, up which the examinee may climb to his highest level without irrelevant hindrances and distractions. Table 9 shows that the French and Spanish tests with which we are concerned approximate a well-built language ladder. There is at least one

TABLE 9

Difficulties of Individual Questions

DISTRIBUTIONS OF QUESTIONS IN EACH PART OF THE FRENCH AND SPANISH TESTS ACCORDING TO PER CENTS OF CORRECT ANSWERS FROM 400 STUDENTS (100 FROM EACH OF THE 4 NORMAL CLASSES, BUT OTHERWISE UNSELECTED). THE FIRST ENTRY IN COLUMN ONE AT THE LEFT SHOWS THAT IN PART I OF THE FRENCH TEST THERE WERE 4 VOCABULARY ITEMS ANSWERED CORRECTLY BY 80 TO 90% OF THE STUDENTS; THE LAST ENTRY IN THAT SAME COLUMN SHOWS THAT THERE WERE 3 VOCABULARY ITEMS CORRECTLY ANSWERED BY LESS THAN 5%, BUT NOT ONE ITEM ANSWERED BY LESS THAN 1.5% OF THE STUDENTS.

PER CENT OF CORRECT ANSWERS	NUMBERS OF QUESTIONS ANSWERED CORRECTLY BY INDICATED PROPORTIONS OF UNSELECTED STUDENTS					
	French Test			Spanish Test		
	Part I	Part II	Part III	Part I	Part II	Part III
90–100%		1		1		
80–89.9%	4	2		6		
70–79.9%	7	2		13	2	1
60–69.9%	9	5	1	7	3	
50–59.9%	13	8	8	8	6	1
40–49.9%	14	15	7	14	15	5
30–39.9%	18	11	2	11	12	7
20–29.9%	13	11	14	10	10	11
10–19.9%	12	5	10	26	10	12
5–9.9%	7		9	3	2	16
1.5–4.9%	3		6	1		9
.5–1.49%			2			
.0–0.5%			1			
Totals	100	60	60	100	60	65

question in each test that was answered correctly by 90% or more of the students, and there were two questions in each test that were answered correctly by not more than 1% of the students. About 14% of all the questions were answered correctly by less than 10% of the students. As was fully expected, Part III in both tests is considerably more difficult than the other Parts. Only one Part III question in each test was answered correctly by more than 60% of the students. But this does not vitiate the validity and reliability of Part III: the important thing is that a test be *smoothly* scaled in difficulty, and while there are no *very* easy items in Part III, they are well-distributed from "60%-correct" degree of difficulty down to less than "1%-correct-answers" degree of difficulty. In this connection, we must also remember that writing is not emphasized as much in the first two semesters as in the last two semesters.

Order of difficulty correlations. — The items of each Part in each test were arranged in orders of difficulty as determined by the per cents of correct responses given by 100 8A, by 100 8B, by 100 9A, and by 100 9B

students. The correlations between two orders of difficulty based upon responses of contiguous classes, e.g. 100 8A and 100 8B students' responses, are consistently higher than the correlations between orders of difficulty based upon responses of e.g. 100 8A and 100 9B students. The correlations between orders of difficulty when each order is based upon responses of 100 students range from 0.60 to 0.95, with an average of about 0.75 for Parts I and II, and about 0.85 for Part III, in both tests. When orders of difficulty are based upon responses of 200 students, the correlations range from 0.87 to 0.96, with an average at about 0.92 for both tests.

If these relationships are confirmed by more extensive investigations, they indicate that modern language learning is highly integrated, or very homogeneous, from the testing viewpoint; the neural bonds that constitute effective language ability are highly interrelated and interdependent. But this evidence may be partially or entirely contradicted by correlations between orders of difficulty based upon responses of two groups of students using different syllabi. On the other hand, it may be that all learning of a given language, regardless of differences between texts and syllabi and methods of instruction used, is characterized in its fundamentals by a fairly large and unvarying core of common neural bonds. In order to enable other users of these tests to make comparisons with orders of difficulty based upon responses of New York City junior high school students, the absolute and relative difficulties of each item in both tests are indicated at the left of each item in columns 1, 2, and 3, on pp. 48 to 85.

Reproduction of the tests.—With the exception of the five [1] columns of figures at the left, which will be explained shortly, and excepting the size of the type and spacing of the lines, the tests are reproduced below exactly as they were administered to New York City junior high school students in June, 1925. Misprints and other errors are faithfully reproduced so that the reader may see exactly what was set before the students. As given to the students, the questions were in 10-point type throughout, with double leads between lines of type in the same question and extra leads between questions. The lines were four inches long, and the whole test was printed in a 12-page booklet, 8½ by 10½. In Part III lines on which the students wrote their responses were 1¾ inches in length, and spaced at least ⅜-inch apart vertically. These tests have been published, and may be secured in quantity, with manuals of directions and norms, from the World Book Co., Yonkers, N. Y. They are called the American Council Beta French and Spanish Tests.

[1] Seven for Part I of the French test. The columns of figures at the left of the test questions will be discussed in detail on pages 86 ff. The tests are reproduced on the following pages:

French: Part I, 48–54; Part II, 55–61; Part III, 62–66.

Spanish: Part I, 67–72; Part II, 73–79; Part III, 80–85

NOTE: *The reader must recall that, with the exception of the columns of italic figures at the left, these tests are reproduced here as they were presented to the students in June, 1925, even to the inclusion of typographical errors, in addition to the linguistic infelicities that escaped our collaborators and proof-readers in that hurried spring of 1925. This is done so that the reader may have a first-hand view of one of the important factors in the experiment, and apprehend from concrete evidence the need for meticulous care in making examinations. These tests have been carefully corrected and revised with the coöperation of several modern language specialists and published as the American Council Beta French and Spanish Tests by the World Book Company, Yonkers, New York.*

New York City Junior High Schools

COMPREHENSIVE FRENCH EXAMINATION: Form A

SCHOOL..... Date, June....., 1925.

Print your full name:.....
(First Name) (Second) Last)

Age.....years.....months. Boy.....Girl.....(check one)

I am just completing French: 8A, 8B, 9A, 9B, RA, RB, RC, RD (circle one)

Name of teacher.....

DIRECTIONS TO STUDENTS: *You will be allowed thirty minutes on each of the three parts of this examination. When you are told to start, turn this page, read the directions for Part I carefully, and answer all the questions that you can in Part I in the time allowed. When the teacher says, "Begin Part II," turn to Part II even if you have not finished Part I. You are not supposed to answer every question; answer only those that you know and don't worry or waste time on the others. If you finish Part I before time is called, look over your work and correct any mistakes you may have made.*

	Score
Part I	
Part II	
Part III	
Total	
Grade	

Copyrighted: Reprinted by Permission

PART I: FRENCH VOCABULARY TEST

DIRECTIONS: *In each line select the English word, or group of words, which is a translation of the French word at the left, and put its number in the parentheses at the right. Do the easy ones first. You will have thirty minutes for this part of the test.*

EXAMPLES:

1. chien 1. chin 2. Chinese 3. dog 4. shine 5. cat.....(3)
2. lire 1. to read 2. lyre 3. to learn 4. lyric 5. to write ...(1)

Per cent right answers (N = 400)	*Order of difficulty within one part*	*Order of difficulty all three parts*	*Index of goodness*	*Order of goodness three parts*	*Henmon frequency*	*Sixteen-book frequency*		
1	2	3	4	5	6	7		
79.8	*5*	*8*	*18*	*180*	*84*	*13*	1. paysan	1. peasant 2. paint 3. country 4. plow 5. safety.....( )
38.8	*50*	*100*	*75*	*16*	*285*	*16*	2. appeler	1. caller 2. appellation 3. learn 4. prelate 5. call.....( )
80.8	*4*	*7*	*41*	*132*	*112*	*14*	3. écouter	1. listen 2. cut 3. acute 4. custom 5. cost.....( )
46.0	*38*	*74*	*53*	*97*	*88*	*16*	4. courir	1. short 2. run 3. road 4. courier 5. course.....( )
88.8	*1*	*3*	*2*	*209*	*85*	*14*	5. argent	1. shine 2. gentle 3. silver 4. eagle 5. urgent .....( )
67.5	*14*	*20*	*64*	*51*	*54*	*15*	6. froid	1. freezing 2. hunger 3. hot 4. cold 5. icy.....( )
31.3	*64*	*124*	*64*	*52*	*52*	*14*	7. rappeler	1. repel 2. hit 3. call back 4. appellation 5. parallel.( )
36.3	*55*	*108*	*65*	*46*	*39*	*13*	8. pont	1. river 2. cross 3. thought 4. bridge 5. boat.....( )
31.8	*63*	*122*	*64*	*53*	*83*	*13*	9. bruit	1. brutal 2. broil 3. brown 4. burn 5. noise.....( )
41.3	*46*	*94*	*28*	*160*	*50*	*14*	10. plaire	1. pleasant 2. please 3. rain 4. complaint 5. plastic..( )
15.3	*84*	*177*	*30*	*156*	*379*	*13*	11. puis	1. since 2. poison 3. then 4. push 5. purse.....( )
32.3	*62*	*121*	*65*	*47*	*176*	*16*	12. lequel	1. which 2. why 3. quell 4. when 5. left.....( )
38.0	*52*	*102*	*77*	*12*	*226*	*16*	13. jusque	1. justice 2. until 3. right 4. jump 5. afterwards....( )
75.5	*7*	*10*	*62*	*60*	*124*	*16*	14. cheval	1. race 2. horse 3. mirror 4. wicked 5. sheep.....( )
68.5	*13*	*18*	*50*	*110*	*423*	*16*	15. aussi	1. thus 2. ass 3. always 4. useful 5. also.....( )

1	2	3	4	5	6	7		
36.8	*53*	*105*	*81*	*7*	*762*	*14*	16. pouvoir	1. pool 2. be able 3. approve 4. poverty 5. duty...( )
59.8	*21*	*32*	*62*	*61*	*50*	*13*	17. neige	1. blanket 2. feathers 3. snow 4. bare 5. white....( )
60.0	*20*	*31*	*58*	*76*	*110*	*13*	18. peur	1. division 2. fear 3. pour 4. bravery 5. poverty....( )
33.5	*60*	*117*	*61*	*65*	*324*	*16*	19. alors	1. finally 2. always 3. then 4. lawful 5. road.....( )
51.8	*29*	*54*	*57*	*82*	*5129*	*15*	20. que	1. so 2. and 3. why 4. that 5. then.....( )
58.3	*22*	*37*	*59*	*73*	*402*	*16*	21. peu	1. odor 2. skin 3. no one 4. lamb 5. little.....( )
45.0	*40*	*80*	*46*	*121*	*38*	*15*	22. déjeuner	1. be youthful 2. drink 3. breakfast 4. get up 5. food.( )
24.3	*75*	*150*	*60*	*70*	*42*	*12*	23. mener	1. menial 2. leader 3. mean 4. meal 5. lead.....( )
39.3	*49*	*99*	*61*	*66*	*90*	*15*	24. asseoir	1. deceive 2. evening 3. handkerchief 4. to seat 5. foolish( )
28.8	*68*	*132*	*47*	*119*	*23*	*12*	25. léger	1. light 2. hallway 3. levity 4. late 5. leisure.....( )
50.5	*31*	*57*	*53*	*98*	*6*	*5*	26. joujou	1. judge 2. jail 3. toy 4. gave 5. torment.....( )
46.8	*36*	*69*	*64*	*54*	*30*	*4*	27. menacer	1. to remind 2. to mince 3. to threaten 4. to warn 5. to mend.....( )
44.3	*43*	*84*	*25*	*169*	..	*8*	28. voyelle	1. young lady 2. vowel 3. fawn 4. violet 5. valley .( )
29.5	*67*	*129*	*13*	*193*	*17*	*5*	29. estimer	1. to pile up 2. to cross 3. to record 4. to stop 5. to appraise.....( )
39.8	*48*	*97*	*17*	*181*	*13*	*7*	30. clef	1. nail 2. key 3. left 4. paw 5. cliff.....( )
15.5	*82*	*174*	*14*	*190*	*16*	*6*	31. vêtement	1. vigorously 2. garment 3. liveliness 4. foot-print 5. vest ( )
76.0	*6*	*9*	*37*	*139*	..	*7*	32. gant	1. gaunt 2. thin 3. glance 4. glove 5. glint.....( )
21.3	*78*	*160*	*14*	*191*	*16*	*7*	33. amitié	1. treaty 2. amicable 3. friendship 4. armistice 5. armed( )
15.0	*85*	*178*	*20*	*177*	..	*7*	34. ruisseau	1. brook 2. nightingale 3. noise 4. roar 5. wine-glass..( )
71.5	*10*	*13*	*0*	*214*	*102*	*5*	35. humain	1. humor 2. hammer 3. humid 4. human 5. worm .( )

1	2	3	4	5	6	7		
61.0	*19*	*29*	*28*	*161*	*25*	*6*	36. pur	1. able 2. put 3. placed 4. impute 5. pure.....( )
74.8	*9*	*12*	*31*	*154*	*9*	*7*	37. féroce	1. iron 2. wild beast 3. fierce 4. precious 5. meager..( )
63.3	*17*	*24*	*34*	*144*	..	*6*	38. bouton	1. bout 2. blow 3. button 4. bottom 5. mountain..( )
28.3	*69*	*133*	*53*	*99*	*14*	*5*	39. projet	1. projection 2. projectile 3. scheme 4. production 5. process.....( )
75.5	*8*	*11*	*−8*	*219*	*8*	*6*	40. miroir	1. mirage 2. lake 3. mirror 4. miracle 5. watch....( )
61.5	*18*	*28*	*68*	*34*	*125*	*12*	41. mort	1. dead 2. morbid 3. empty 4. moth 5. mouth.....( )
66.0	*15*	*21*	*4*	*207*	*68*	*14*	42. fermer	1. ferment 2. strength 3. fence 4. close 5. cloister...( )
81.5	*3*	*6*	*56*	*86*	*5*	*14*	43. samedi	1. sample 2. cereal 3. summer 4. monthly 5. Saturday ( )
51.8	*30*	*55*	*63*	*59*	*30*	*12*	44. lune	1. alone 2. moon 3. lurid 4. shining 5. one.....( )
69.3	*12*	*17*	*70*	*30*	*160*	*16*	45. perdre	1. perdition 2. urge 3. poor 4. lose 5. purge.....( )
53.0	*27*	*49*	*61*	*67*	*26*	*14*	46. mer	1. sickness 2. sea 3. mother 4. evil 5. mirror.....( )
50.0	*33*	*60*	*62*	*62*	*1540*	*15*	47. on	1. one 2. ass 3. into 4. them 5. door.....( )
37.8	*57*	*111*	*52*	*103*	*55*	*15*	48. remercier	1. thank 2 pity 3. remit 4. gift 5. reward.....( )
30.8	*65*	*125*	*74*	*19*	*68*	*10*	49. gauche	1. right 2. station 3. coach 4. left 5. behind.....( )
22.0	*77*	*157*	*31*	*155*	*77*	*12*	50. vent	1. opening 2. vengeance 3. cloudy 4. vanquish 5. wind ( )
54.5	*25*	*47*	*55*	*88*	*150*	*16*	51. attendre	1. listen 2. attain 3. waste 4. await 5. waiter.....( )
33.3	*61*	*118*	*70*	*31*	*20*	*12*	52. tôt	1. total 2. carry 3. top 4. soon 5. when.....( )
44.8	*41*	*81*	*58*	*77*	*56*	*13*	53. tromper	1. deceive 2. break 3. trumpet 4. romp 5. thunder..( )
18.8	*80*	*167*	*51*	*106*	*25*	*11*	54. moitié	1. morsel 2. half 3. moat 4. mortal 5. box.....( )
83.5	*2*	*5*	*48*	*116*	*10*	*13*	55. lundi	1. Monday 2. tunnel 3. mud 4. under 5. Tuesday..( )

1	2	3	4	5	6	7		
47.5	*35*	*68*	*84*	*3*	*134*	*15*	56. lever	1. never 2. light 3. raise 4. levity 5. let fall.....( )
28.8	*70*	*134*	*61*	*68*	*68*	*14*	57. champ	1. field 2. bite 3. grass 4. chant 5. chance.....( )
46.3	*37*	*71*	*79*	*9*	*151*	*15*	58. droit	1. lawful 2. drought 3. narrow 4. way 5. right.....( )
57.0	*23*	*39*	*54*	*92*	*41*	*12*	59. été	1. spring 2. eaten 3. either 4. out 5. summer.....( )
49.5	*34*	*62*	*86*	*2*	*14*	*12*	60. magasin	1. magistrate 2. material 3. magician 4. store 5. manager.....( )
44.5	*42*	*82*	*73*	*21*	*78*	*14*	61. soleil	1. round 2. shining 3. yellow 4. solid 5. sun.....( )
52.8	*28*	*50*	*71*	*27*	*14*	*12*	62. bateau	1. boat 2. batter 3. gift 4. battle 5. wreath.....( )
44.3	*44*	*85*	*54*	*93*	*105*	*13*	63. autant	1. author 2. always 3. around 4. as much 5. aunt...( )
19.5	*79*	*166*	*36*	*142*	*57*	*15*	64. oreille	1. really 2. oriole 3. window 4. ear 5. oracle.....( )
10.8	*90*	*191*	*27*	*162*	*327*	*16*	65. falloir	1. do without 2. fallacy 3. fatal 4. be necessary 5. fault( )
25.3	*74*	*146*	*39*	*136*	..	*4*	66. impur	1. foul 2. impish 3. empire 4. impute 5. to impede..( )
29.8	*66*	*128*	*65*	*48*	*164*	*8*	67. pas	1. pass 2. past 3. foot 4. step 5. goat.....( )
63.5	*16*	*23*	*−2*	*216*	*10*	*5*	68. poudre	1. dust 2. powder 3. power 4. chicken 5. to thrust.( )
34.3	*59*	*114*	*43*	*130*	*23*	*8*	69. confier	1. to trust 2. to try 3. to tell 4. to recite 5. confidence ( )
27.3	*71*	*135*	*35*	*143*	*13*	*7*	70. queue	1. cue 2. curse 3. queer 4. quick 5. tail.....( )
38.5	*51*	*101*	*19*	*179*	*9*	*4*	71. proclamer	1. to exclaim 2. to announce 3. town-crier 4. to rejoice 5. to promise.....( )
70.8	*11*	*15*	*8*	*204*	..	*4*	72. ballon	1. balloon 2. ball 3. feat 4. beacon light 5. lighthouse ( )
12.5	*89*	*187*	*27*	*163*	*36*	*8*	73. fauteuil	1. fault 2. mistake 3. tall clock 4. armchair 5. townhall( )
18.5	*81*	*169*	*21*	*176*	*6*	*7*	74. haïr	1. hair 2. to hate 3. to breathe 4. to hasten 5. to flee.( )
15.5	*83*	*175*	*33*	*148*	*23*	*8*	75. précipiter	1. to cast 2. precipice 3. to give away 4. steep 5. precious( )

1	2	3	4	5	6	7			
25.5	*72*	*142*	*29*	*158*	*8*	*4*	76. étoffe	1. attack 2. stuff 3. candy 4. crown 5. stage	( )
22.5	*76*	*155*	*32*	*139*	..	*4*	77. sabot	1. cabbage 2. sword 3. wooden shoe 4. wiseacre 5. row-boat	( )
6.3	*96*	*204*	*11*	*197*	*27*	*6*	78. augmenter	1. argument 2. to persuade 3. to praise 4. to amount 5. to enlarge	( )
5.0	*97*	*208*	*12*	*194*	..	*6*	79. mûrir	1. wall 2. to move 3. to mutilate 4. to ripen 5. to murmur	( )
13.8	*86*	*181*	*23*	*171*	*16*	*8*	80. gras	1. fat 2. grass 3. large 4. gray 5. grape	( )
54.3	*26*	*48*	*57*	*83*	..	*6*	81. après-demain	1. afternoon 2. behind 3. yesterday 4. day-after-tomorrow 5. next week	( )
14.3	*87*	*183*	*29*	*159*	*35*	*8*	82. sombre	1. shady 2. sleepy 3. sorry 4. large hat 5. dark	( )
35.0	*58*	*112*	*10*	*199*	..	*7*	83. cadet	1. frame 2. clasp 3. junior 4. gift 5. cake	( )
36.5	*54*	*107*	*53*	*100*	*24*	*6*	84. mien	1. mien 2. more 3. aspect 4. mine 5. opening	( )
45.3	*39*	*78*	*−4*	*218*	..	*5*	85. enseigne	1. lord 2. teacher 3. notebook 4. meaning 5. sign	( )
55.5	*24*	*43*	*16*	*183*	*10*	*5*	86. signer	1. to sign 2. to restore 3. to signal 4. to adopt 5. to seal	( )
13.0	*88*	*186*	*27*	*164*	*10*	*5*	87. souhait	1. wish 2. wind 3. south 4. whistle 5. folksong	( )
40.3	*47*	*96*	*2*	*210*	*7*	*6*	88. romain	1. romance 2. Roman 3. roamer 4. robin 5. lettuce	( )
9.0	*92*	*197*	*26*	*167*	*12*	*8*	89. envers	1. against 2. inverse 3. envoy 4. towards 5. poetry	( )
44.3	*45*	*86*	*32*	*152*	..	*4*	90. couche	1. crutch 2. basket 3. bed 4. hide 5. blanket	( )
6.8	*95*	*201*	*−3*	*217*	..	*4*	91. loge	1. box 2. log 3. load 4. lobby 5. logical	( )
9.5	*91*	*194*	*20*	*178*	*16*	*6*	92. espoir	1. despair 2. spirit 3. hope 4. spy 5. pomp	( )
3.0	*100*	*214*	*7*	*206*	*10*	*7*	93. renoncer	1. to renew 2. renown 3. to adjure 4. to abhor 5. reliance	( )

1	2	3	4	5	6	7		
35.5	*56*	*109*	*−8*	*220*	*7*	*6*	94. démontrer	1. to cast aside 2. to dismount 3. watch 4. demonstrate 5. to look down upon.....( )
6.3	*94*	*200*	*15*	*189*	..	*4*	95. marée	1. sea 2. tide 3. husband 4. marriage 5. march....( )
9.0	*93*	*198*	*10*	*200*	*28*	*5*	96. coq	1. cork 2. barrel 3. chick 4. rooster 5. crack.....( )
50.3	*32*	*58*	*12*	*195*	*17*	*5*	97. régler	1. to record 2. to glance 3. to regulate 4. to reform 5. to reign.....( )
25.5	73	*143*	*22*	*174*	..	*4*	98. grille	1. dining room 2. grate 3. a tall animal 4. graceful 5. slender.....( )
4.0	*99*	*212*	*14*	*192*	*28*	*7*	99. ému	1. African animal 2. emulated. 3. emit 4. thirst 5. agitated.....( )
4.5	*98*	*209*	*17*	*182*	*6*	*5*	100. dessein	1. dessert 2. underneath 3. plan 4. drapery 5. destiny.( )

Per cent right answers (N=400)	Order of difficulty within one part	Order of difficulty all three parts	Index of goodness	Order of goodness three parts
1	2	3	4	5
91.5	*1*	*1*	*4*	*208*
89.8	*2*	*2*	*16*	*184*
68.0	*6*	*19*	*41*	*133*
70.5	*5*	*16*	*50*	*111*
44.3	*27*	*87*	*83*	*5*
57.5	*12*	*41*	*68*	*35*
63.3	*7*	*25*	*71*	*28*
88.5	*3*	*4*	*16*	*185*

PART II: FRENCH COMPREHENSION TEST

DIRECTIONS: *In each of the following statements there are five alternatives. Select the one which makes the statement most reasonable and put its number in the parenthesis at the right, as in the examples. You have thirty minutes for this part of the examination.*

EXAMPLES:

1. Le chien est — 1. un livre 2. un crayon 3. un animal 4. un ours 5. un sac ....(3)
2. On se sert d'encre pour — 1. manger 2. boire 3. courir 4. nager 5. écrire ....(5)

1. La capitale de la France est — 1. New York 2. Londres 3. le président 4. Paris 5. la Seine.....( )
2. La langue officielle des États-Unis est—1. l'espagnol 2. le français 3. l'anglais 4. impossible 5. très mauvaise.....( )
3. La pomme est — 1. un animal 2. un métal 3. un légume 4. une mouche 5. un fruit.....( )
4. Comme sa mère était très malade, elle a fait chercher — 1. le capitaine 2. le médecin 3. la couturière 4. le maître 5. le journal.....( )
5. Quand on fait une promenade au grand soleil, le chapeau se porte sur — 1. le nez 2. le pied 3. la chaise 4. la tête 5. le genou.....( )
6. La neige est — 1. verte 2. blanche 3. chaude 4. inutile 5. bleue.....( )
7. Quand quelqu'un pose une question, on doit presque toujours — 1. partir 2. chanter 3. se coucher 4. prier 5. répondre.....( )
8. Jean a passé une année entière à Paris. Il a appris à parler la langue du pays. La langue qu'il a apprise est — 1. l'anglais 2. l'espagnol 3. l'italien 4. le français 5. l'allemand.....( )

1	2	3	4	5
42.5	*32*	*92*	*68*	*36*
52.0	*16*	*51*	*68*	*37*
51.8	*18*	*56*	*60*	*71*
55.3	*14*	*44*	*54*	*94*
48.0	*19*	*65*	*88*	*1*
71.0	*4*	*14*	*38*	*138*
46.3	*22*	*72*	*78*	*10*
59.3	*11*	*33*	*75*	*17*
62.8	*8*	*26*	*67*	*41*
45.8	*25*	*76*	*49*	*112*
60.5	*10*	*30*	*70*	*32*

9. Quand j'aurai beaucoup d'argent j'en donnerai aux pauvres; donc je serai — 1. avare. 2. faible 3. dangereux 4. généreux 5. paresseux.....( )

10. Un enfant poli, quand on lui donne quelquechose, dit toujours — 1. bonjour 2 au revoir 3. merci 4. comment 5. de bonne heure.....( )

11. De quelle couleur est le ciel quand il fait beau? — 1. noir 2. cher 3. heureux 4. blanc 5. bleu.....( )

12. Ma soeur a été bien malade mais maintenant elle se porte très — 1. mal 2. triste 3. bien 4. mauvais 5. première.....( )

13. Quand on a faim, on veut — 1. manger 2. a froid 3. est malade 4. chanter 5. a soif.....( )

14. Qu'est-ce qui sépare l'Amérique du Nord de l'Europe? — 1. la France 2. l'Angleterre 3. l'Atlantique 4. la forêt 5. la rue.....( )

15. Le dimanche est un jour — 1. de travail 2. de fête 3. froid 4. long 5. en hiver.( )

16. Pour apprendre à lire et à écrire il faut aller — 1. à la forêt 2. au théâtre 3. en automobile 4. à l'école 5. chez le dentiste.....( )

17. Quand on a mal aux dents il faut aller chez — 1. le fermier 2. le docteur 3. le tailleur 4. la couturière 5. le dentiste.....( )

18. Les lits sont dans — 1. le salon 2. la chambre à coucher 3. la cuisine 4. la salle à manger 5. la porte.....( )

19. Nous irons chercher des livres pour étudier nos leçons d'histoire — 1. à la campagne 2. au théâtre 3. au jardin 4. dans la bibliothèque 5. dans la mer.( )

1	2	3	4	5	
36.8	*36*	*106*	*74*	*20*	20. Quand le maître est entré dans la salle de classe un élève écrivait au tableau noir; cet élève tenait à la main gauche un livre, et dans l'autre — 1. un fil 2. un morceau de craie 3. un fusil 4. un crayon 5. un canif.....( )
47.8	*20*	*67*	*66*	*43*	21. Quand Marie s'addresse à la soeur de sa mère elle l'appelle — 1. mère 2. soeur 3. nièce 4. tante 5. voisine.....( )
30.3	*44*	*127*	*66*	*44*	22. S'il pleut quand vous voudrez sortir, prenez — 1. une canne 2. un livre 3. un parapluie 4. une chaise 5. un journal.....( )
46.3	*21*	*70*	*68*	*38*	23. Nous étions à New York pour assister à une des fêtes le plus importantes des États-Unis; nous ne nous sommes pas amusés parce qu'il faisait si chaud. Quelle était la date de cette fête? — 1. le vingt-cinq décembre 2. le premier janvier. 3. le quatre juillet 4. le vingt et un février 5. le cinq novembre..( )
34.3	*39*	*115*	*64*	*55*	24. Au printemps les arbres se couvrent — 1. de beurre 2. de feuilles 3. de soie 4. de couteaux 5. d'or.....( )
57.3	*13*	*38*	*68*	*39*	25. Nous avons appris par coeur, "Le Corbeau et le Renard." Le renard est — 1. un animal 2. une fleur 3. un fruit 4. un chien 5. un oiseau.....( )
52.0	*17*	*52*	*84*	*4*	26. En été, les feuilles de presque tous les arbres sont — 1. vertes 2. petites 3. sèches 4. rouges 5. rien.....( )
44.0	*28*	*88*	*53*	*101*	27. Robert a douze ans, c'est-à-dire, trois ans de plus que sa soeur, Jeanne. Quel âge Jeanne a-t-elle? — 1. quinze ans 2. neuf ans 3. neuf mois 4. huit semaines 5. quelques jours.....( )
31.5	*42*	*123*	*72*	*23*	28. Un bonne qui passe sa vie sur un bateau s'appelle — 1. un voisin 2. un matelot 3. un médecin 4. un oncle 5. un neveu.....( )

1	2	3	4	5	
39.5	*34*	*98*	*72*	*24*	29. Les enfants qui ne demeurent pas loin viennent à l'école — 1. en auto 2. en voiture 3. à cheval 4. en omnibus 5. à pied.....( )
22.8	*57*	*154*	*33*	*149*	30. Aux murs de la salle de classe sont suspendus — 1. des lits 2. des chaises 3. des tableaux 4. des livres 5. des mouchoirs.....( )
40.5	*33*	*95*	*82*	*6*	31. Quel est le pluriel du mot "l'oeil"? — 1. les yeux 2. les oreilles 3. les bouches 4. les ongles 5. des cheveux.....( )
35.5	*37*	*110*	*54*	*95*	32. Pour manger la soupe il faut se servir — 1. d'un couteau 2. d'un crayon 3. du papier 4. d'un canif 5. d'une cuiller.....( )
21.0	*53*	*162*	*22*	*175*	33. Notre vache nous fournit autant qu'il nous faut — 1. d'eau 2. de vin 3. de cidre 4. de lait 5. de farine.....( )
44.5	*26*	*83*	*49*	*113*	34. A la fin d'une phrase qui pose une question, on met — 1. une virgule 2. un point 3. un point d'interrogation 4. un verbe 5. une fois.....( )
62.0	*9*	*27*	*59*	*74*	35. Une personne qui n'a pas assez d'argent pour acheter du pain, est — 1. riche 2. contente 3. pauvre 4. de bonne humeur 5. célèbre.....( )
30.5	*43*	*126*	*51*	*107*	36. La langue est une partie — 1. de l'oreille 2. de la bouche 3. de la boîte 4. de la maison 5. de la chambre.....( )
32.5	*41*	*120*	*37*	*140*	37. Un paysan est un homme qui travaille — 1. dans l'eau 2. dans l'air 3. à la campagne 4. dans les arbres 5. sous la terre.....( )
43.0	*31*	*91*	*65*	*49*	38. Quand un monsieur ne se sert pas de son mouchoir, il le tient — 1. dans son soulier 2. sur la fenêtre 3. sous la terre 4. dans sa poche 5. sur sa tête...( )

1	2	3	4	5	
46.0	*24*	*75*	*76*	*15*	39. Ce jeune soldat a perdu le bras droit à la Grande Guerre; quel bras lui reste-t-il? — 1. celui qu'il a perdu 2. les yeux 3. le fusil 4. le pied 5. le bras gauche.....( )
18.5	*57*	*170*	*49*	*114*	40. Les personnes à la campagne qui se lèvent de bonne heure voient le magnifique spectacle du soleil qui se lève — 1. à midi 2. à l'est 3. au nord 4. à l'ouest 5. à minuit.....( )
24.3	*50*	*151*	*54*	*96*	41. Les doigts de cette grande dame sont couverts de — 1. bagues 2. feuilles 3. lait 4. bois 5. fer.....( )
43.3	*30*	*90*	*55*	*89*	42. Les grands auteurs qui ont fait la gloire de la littérature française se sont servis d'encre pour — 1. penser 2. marcher 3. écrire 4. danser 5. chanter( )
33.0	*40*	*119*	*62*	*63*	43. Ma tante ira chez la modiste pour se faire faire un — 1. habit 2. ruban 3. chapeau 4. livre 5. gant.....( )
25.8	*48*	*140*	*44*	*128*	44. C'est le matin. Il est onze heures et demie. Paul va partir dans vingt minutes. A quelle heune partira-t-il? — 1. à minuit 2. à onze heures vingt 3. à midi moins dix 4. à une heure 5. à l'école.....( )
29.0	*45*	*131*	*61*	*69*	45. J'achèterai une douzaine d'oeufs; donc j'en aurai — 1. trois 2. douzième 3. douze 4. six 5. une dizaine.....( )
18.8	*56*	*168*	*45*	*125*	46. Tout le monde sait que la partie la plus haute d'une maison est — 1. l'escalier 2. la porte 3. le tiroir 4. le salon 5. le toit.....( )
35.0	*38*	*113*	*60*	*72*	47. A table on boit de l'eau — 1. d'un gant 2. dans un verre 3. avec un panier 4. d'un ruban 5. avec les yeux.....( )

1	2	3	4	5	
26.5	*46*	*137*	*48*	*117*	48. La laine nous vient — 1. du cheval 2. du rat 3. de la cigale 4. du mouton 5. du chien.....( )
46.3	*23*	*73*	*73*	*22*	49. Les enfants très jeunes aiment beaucoup les — 1. joujoux 2. idées 3. travaux 4. lectures difficiles 5. pierres précieuses.....( )
44.0	*29*	*89*	*64*	*56*	50. Quand un ami vient vous rendre visite chez vous, il est d'usage de lui offrir — 1. un soulier 2. une chaise 3. une lettre 4. un drapeau 5. un citron.....( )
26.5	*41*	*138*	*51*	*108*	51. Elle allait tous les matins chez le boulanger chercher — 1. de la viande 2. du chocolat 3. des bonbons 4. du pain 5. du sel.....( )
37.5	*35*	*103*	*75*	*18*	52. Quand on se dépêche afin d'arriver à la gare à temps, on marche — 1. lentement 2. à meilleur marché 3. surtout 4. vite 5. lent.....( )
21.5	*52*	*159*	*46*	*122*	53. Les enfants joueront en été sur la plage dans — 1. la neige 2. le sable 3. la farine 4. le sucre 5. la pierre.....( )
55.3	*15*	*45*	*77*	*13*	54. Je vois dans une cage dorée quelquechose qui a deux petits pieds et deux petites ailes; c'est — 1. un chat 2. un oiseau 3. un soldat 4. un nuage 5. un singe..( )
18.5	*58*	*171*	*51*	*109*	55. Il était debout dans le salon près du piano; son petit ami était assis à ses pieds sur — 1. le tapis 2. le quai 3. le piano 4. le sable 5. le mur.....( )
20.8	*54*	*163*	*48*	*118*	56. Un homme riche qui ne donne jamais aux pauvres et qui se passe des choses qui lui sont nécessaires pour le bonheur et pour la santé est — 1. généreux 2. avare 3. très bon envers les pauvres 4. bien aimé 5. très sage.....( )
13.3	*60*	*184*	*26*	*168*	57. Il n'y a rien dans la bouteille; donc elle est — 1. remplie 2. froide 3. haute 4. vide 5. lourde.....( )

1	2	3	4	5	
25.5	*49*	*144*	*52*	*104*	58. L'aiguille est un instrument qui s'emploie beaucoup par — 1. les rois 2. les couturières 3. les souris 4. les singes 5. les étrangers.....( )
20.3	*55*	*164*	*27*	*165*	59. Parmi des voleurs l'homme honnête est — 1. heureux 2. sans peur 3. très content 4. en danger 5. sauf.....( )
15.0	*59*	*179*	*44*	*128*	60. Pour faire son travail la couturière a besoin — 1. d'un cahier 2. d'un dé 3. d'un soulier 4. des nids 5. d'un chapeau.....( )

Per cent right answers (N-400)	*Order of difficulty within one part*	*Order of difficulty all three parts*	*Index of goodness*	*Order of goodness three parts*
1	2	3	4	5
64.5	*1*	*22*	*71*	*29*
57.8	*3*	*35*	*68*	*40*
49.8	*10*	*61*	*37*	*141*
55.3	*7*	*46*	*43*	*131*
25.8	*22*	*141*	*66*	*45*
56.3	*5*	*41*	*58*	*78*
29.5	*19*	*130*	*80*	*8*
25.0	*24*	*147*	*72*	*25*

PART III: FRENCH GRAMMAR COMPLETION

DIRECTIONS: *Each of the English sentences below is followed by a translation which is correct as far as it goes, but which is incomplete. Each translation can be completed correctly by inserting one, two or three words in each blank (——). When you have decided what is necessary to complete the translation, write it on the line at the right of the translation. WRITE LEGIBLY. DO NOT SPEND TOO MUCH TIME on any one sentence. Do all the easy ones first, then go back to the harder ones if you have time. You will have thirty minutes for this part of the examination. Write carefully and clearly, putting your answers at the right as in the examples.*

EXAMPLES:

1. I see some dogs. Je vois (——) chiens..... *des*
2. I am hungry. (——) faim ..... *J'ai*
3. I make him read this book. Je lui (——) ce livre..... *fais lire*

1. Give me the pen-knife. (——) le canif.....
2. Her dress is white. Sa robe (——).....
3. She has her paper. Elle a (——) papier.....
4. My name is Louise. Je (——) Louise.....
5. Bring me the others. Apportez moi (——).....
6. There is a beautiful house. Voilà une (——) maison.....
7. I eat some bread with each meal. Je mange (——) à chaque repas.....
8. He gives me the papers. (——) les papiers.....

1	2	3	4	5	
57.0	*4*	*40*	*34*	*145*	9. She found the boy's pencil. Elle a trouvé le crayon (——) garçon
37.5	*17*	*104*	*78*	*11*	10. It is Mary's hat. C'est (——) Marie
41.8	*16*	*93*	*45*	*126*	11. I speak to the children. Je parle (——) enfants
59.0	*2*	*34*	*32*	*153*	12. I have taken my ink. J'ai pris (——)
48.8	*12*	*64*	*55*	*90*	13. These flowers are beautiful. (——) sont jolies
34.3	*18*	*116*	*72*	*26*	14. Three days ago. (——) trois jours
23.0	*28*	*153*	*64*	*57*	15. We were with him. Nous étions (——)
45.8	*14*	*77*	*49*	*115*	16. Speak to the man. Parlez (——)
22.0	*30*	*158*	*67*	*42*	17. I have my knife and he has his. J'ai mon canif et il a (——)
13.3	*38*	*185*	*53*	*102*	18. She brings them to me. Elle (——) apporte
50.3	*9*	*40*	*64*	*58*	19. Put it on the table. Posez-le (——) table
48.0	*13*	*66*	*46*	*123*	20. There are forty-three students in that class. Il y a (——) élèves dans cette classe
52.0	*8*	*53*	*70*	*33*	21. The weather is fine. (——)

1	2	3	4	5	
21.3	*31*	*161*	*58*	*79*	22. Which pen do you wish? (——) voulez-vous?
24.5	*25*	*148*	*55*	*91*	23. They have eighty-five books. Ils ont (——) livres.
25.5	*23*	*145*	*59*	*75*	24. Because he was ill. (——) était malade.
22.5	*29*	*156*	*57*	*84*	25. He works every day. Il travaille (——)
17.8	*33*	*172*	*57*	*85*	26. We have ours. Nous avons (——)
27.0	*20*	*136*	*52*	*105*	27. I do not have any money. Je n'ai pas (——)
26.3	*21*	*139*	*65*	*50*	28. Do you live in the green house? Demeurez-vous dans (——)?
23.3	*27*	*152*	*62*	*64*	29. What do you say? (——)?
9.5	*44*	*195*	*34*	*146*	30. She has her pen and that of her friend. Elle a sa plume et (——) de son amie
9.8	*43*	*193*	*30*	*157*	31. Here are two books; don't take this one. Voici deux livres; ne prenez pas (——)
24.5	*26*	*149*	*77*	*14*	32. Those pencils are mine. Ces plumes sont (——)
13.8	*37*	*182*	*40*	*134*	33. Take the other one. Prenez (——)
9.5	*45*	*196*	*33*	*150*	34. The penknife that is on the table is small. Le canif (——) sur la table est petit

1	2	3	4	5	
55.8	*6*	*42*	*47*	*120*	35. Come quickly! Venez (——)!
10.0	*42*	*192*	*23*	*172*	36. He does not eat much. Il ne mange (——).
12.3	*39*	*188*	*45*	*127*	37. We were going home. Nous (——) chez nous.
14.5	*36*	*180*	*39*	*137*	38. They are calling Jacques. Ils (——) Jacques
49.0	*11*	*63*	*58*	*80*	39. How old are you? (——) âge avez-vous?
6.5	*48*	*203*	*23*	*173*	40. I looked for my pens but I did not find them. J'ai cherché mes plumes, mais je ne les ai pas (——)
8.0	*46*	*199*	*27*	*166*	41. Will you not be on time? (——) à temps?
20.3	*32*	*165*	*58*	*81*	42. I am sure that you are wrong. Je suis sûr que vous (——).
17.0	*34*	*173*	*56*	*87*	43. They left early. Ils partirent (——).
45.3	*15*	*79*	*16*	*186*	44. I am well. Je (——)
12.3	*40*	*189*	*34*	*147*	45. Did they sell their house? (——) leur maison?
15.5	*35*	*176*	*46*	*124*	46. We go there frequently. Nous y allons (——)
5.8	*50*	*206*	*12*	*196*	47. We were there in the springtime. Nous étions là (——)

1	2	3	4	5	
12.0	*41*	*190*	*40*	*135*	48. She has been here since this morning. Elle est ici (——) ce matin.....
4.3	*52*	*210*	*10*	*201*	49. She has some beautiful gloves. Elle a (——) gants.....
5.5	*51*	*207*	*16*	*187*	50. I shall go at once. J'irai (——).....
2.0	*57*	*217*	*8*	*205*	51. These two boys were born on the same day. Ces deux garçons (——) le même jour.....
3.8	*54*	*213*	*11*	*198*	52. I shall speak to him when he comes. Je lui parlerai quand il (——).....
4.3	*53*	*211*	*2*	*211*	53. We go to my aunt's on Wednesdays. Nous allons chez ma tante (——).....
6.8	*47*	*202*	*24*	*170*	54. I bought them at a bargain. Je les ai achetés (——).....
6.0	*49*	*205*	*16*	*188*	55. The weather is hot in July. Il fait (——).....
1.0	*58*	*218*	*2*	*212*	56. I am fifteen years old. Je suis âgé (——).....
.0	*60*	*220*	*0*	*215*	57. As for you, go! (——) vous, allez-vous-en!.....
2.5	*55*	*215*	*10*	*202*	58. He must work. Il (——) travailler.....
.8	*59*	*219*	*2*	*213*	59. We are upstairs. Nous sommes (——).....
2.5	*56*	*216*	*10*	*203*	60. He was among them. Il était (——).....

New York City Junior High Schools

COMPREHENSIVE SPANISH EXAMINATION: Form A

School..... Date, June....., 1925.

Print your name in full.....
(First Name) (Second) (Last)

Age.....years.....months. Boy.....Girl.....(check one)

I am just completing Spanish 8A, 8B, 9A, 9B, RA, RB, RC, RD (circle one)

Name of teacher.....

Directions to Students: *You will be allowed thirty minutes on each of the three parts of this examination. When you are told to start, turn this page, read the directions for Part I carefully, and answer all the questions that you can in Part I in the time allowed. When the teacher says, "Begin Part II," turn to Part II even if you have not finished Part I. You are not supposed to be able to answer every question; answer only those that you know and don't worry or waste time on the others. If you finish Part I before time is called, look over your work and correct any mistakes you may have made.*

	Score
Part I	
Part II	
Part III	
Total	
Grade	

Copyrighted: Reprinted by Permission

PART I: SPANISH VOCABULARY

DIRECTIONS: *In each line select the English word or phrase which is a translation of the Spanish word at the left, and put its number in the parentheses at the right, as in the examples. Do all the easy ones first. Thirty minutes.*

EXAMPLES:

libro 1. paper 2. book 3. pencil 4. free 5. make.....(2)
papel 1. popular 2. read 3. fly 4. pray 5. paper.....(5)
escribir 1. scribe 2. play 3. write 4. pen 5. ink.....(3)

Per cent right answers (N = 400)	*Order of difficulty within one part*	*Order of difficulty all three parts*	*Index of goodness*	*Order of goodness whole test (three parts)*		
1	2	3	4	5		
89.0	*2*	*2*	*5*	*197*	1. zapato	1. leather 2. spade 3. shoe 4. claw 5. strike.....( )
63.8	*26*	*30*	*55*	*31*	2. querer	1. make quiet 2. wish 3. silence 4. burn 5. remove.....( )
64.0	*25*	*29*	*37*	*80*	3. mal	1. badly 2. sea 3. male 4. mole 5. crime.....( )
58.5	*28*	*35*	*52*	*41*	4. desayuno	1. breakfast 2. disinfect 3. desert 4. broken up 5. twilight..( )
71.0	*18*	*21*	*13*	*170*	5. primavera	1. spring 2. winter 3. principal 4. first 5. manner.....( )
79.8	*8*	*8*	*−1*	*216*	6. página	1. herald 2. page 3. little 4. pigeon 5. preface.....( )
56.0	*30*	*37*	*47*	*56*	7. plata	1. silver 2. plot 3. braid 4. worse 5. prize.....( )
76.8	*11*	*11*	*12*	*173*	8. tocar	1. join 2. touch 3. trick 4. attempt 5. trouble.....( )
34.8	*54*	*96*	*42*	*68*	9. salir	1. sail 2. save 3. know 4. walk 5. go out.....( )
82.8	*5*	*5*	*1*	*211*	10. botella	1. glass 2. round 3. mirror 4. bottle 5. fight.....( )
81.8	*6*	*6*	*28*	*105*	11. árbol	1. raise 2. wooden bowl 3. arbor 4. tree 5. leaf.....( )
52.0	*34*	*44*	*3*	*204*	12. cerrar	1. inclose 2. seal 3. shut 4. sift 5. cease.....( )
84.8	*4*	*4*	*7*	*192*	13. beber	1. fondle 2. wash 3. drink 4. weak 5. baby.....( )
81.0	*7*	*7*	*16*	*158*	14. corbata	1. timid 2. necktie 3. comb 4. coal 5. heating.....( )
47.3	*36*	*54*	*17*	*154*	15. mientras	1. many 2. even 3. same 4. whole 5. while.....( )

1	2	3	4	5			
45.8	*39*	*59*	*55*	*32*	16. poder	1. powder 2. put 3. slow 4. prove 4. be able 5. own....( )
76.5	*12*	*12*	*5*	*198*	17. detrás	1. before 2. behind 3. below 4. across 5. moreover.....( )
72.3	*14*	*17*	*3*	*205*	18. techo	1. roof 2. touch 3. cotton 4. narrow 5. taught.....( )
44.5	*42*	*66*	*−5*	*221*	19. regla	1. jewel 2. watch 3. lake 4. gift 5. rule.....( )
20.5	*69*	*136*	*16*	*156*	20. noticia	1. billboard 2. enticement 3. news 4. notices 5. table of contents( )
70.8	*19*	*22*	*4*	*202*	21. naranja	1. apple 2. sail 3. orange tree 4. orange 5. blanket.....( )
59.5	*29*	*48*	*19*	*144*	22. pálido	1. faint 2. pale 3. shovel 4. paladin 5. white.....( )
68.5	*22*	*25*	*34*	*88*	23. cuatrocientos	1. correspondence 2. string quartet 3. forty 4. knowledge 5. four hundred.....( )
92.5	*1*	*24*	*−16*	*224*	24. ocho	1. eight 2. eye 3. order 4. another 5. ear.....( )
45.8	*40*	*60*	*31*	*140*	25. cuello	1. cheek 2. cruller 3. footstep 4. neck 5. quail.....( )
43.3	*47*	*75*	*40*	*75*	26. pintura	1. paper of pins 2. pincers 3. painting 4. apron 5. landscape( )
87.0	*3*	*3*	*−1*	*217*	27. siete	1. always 2. seat 3. century 4. seven 5. siesta.....( )
52.5	*33*	*43*	*29*	*99*	28. comprador	1. thief 2. compare 3. communication 4. bearer 5. buyer...( )
50.3	*35*	*48*	*23*	*127*	29. mediodía	1. mood 2. noon 3. mediate 4. half-time 5. midnight.....( )
17.8	*76*	*155*	*48*	*53*	30. lástima	1. pity 2. lately 3. last one 4. tear 5. enduring.....( )
43.5	*46*	*73*	*20*	*140*	31. obscuro	1. obstinate 2. dangerous 3. blunt 4. dark 5. obstruction..( )
68.0	*23*	*26*	*34*	*89*	32. estación	1. season 2. sojourn 3. stability 4. explosion 5. detention..( )
70.5	*20*	*23*	*34*	*90*	33. estómago	1. tomato 2. stranger 3. taken out 4. foreign 5. stomach...( )
33.0	*58*	*101*	*27*	*110*	34. salud	1. salad 2. well 3. celery 4. health 5. salary.....( )
71.8	*15*	*18*	*23*	*126*	35. vendedor	1. buyer 2. market 3. conqueror 4. business 5. seller.....( )

1	2	3	4	5			
71.8	*16*	*19*	*3*	*206*	36. julio	1. pretty 2. June 3. young 4. Julius 5. July	( )
44.0	*45*	*72*	*23*	*128*	37. tratar	1. retard 2. take a trip 3. trace 4. treat 5. swallow	( )
31.5	*60*	*108*	*29*	*100*	38. desde	1. out of 2. from 3. after 4. frown 5. sense	( )
29.3	*61*	*113*	*1*	*212*	39. mercado	1. market 2. marked 3. label 4. food supply 5. moored	( )
71.5	*17*	*20*	*20*	*136*	40. setenta	1. sixty 2. week 3. seven 4. suddenly 5. seventy	( )
22.5	*65*	*131*	*49*	*49*	41. valer	1. to be brave 2. tolerate 3. to be worth 4. value 5. sail	( )
20.8	*68*	*137*	*18*	*149*	42. firmar	1. sign 2. ascertain 3. fix 4. tighten 5. company	( )
42.5	*48*	*78*	*11*	*176*	43. sonido	1. exit 2. health 3. solid 4. sound 5. drowned	( )
35.3	*53*	*95*	*8*	*189*	44. bello	1. roar 2. bear 3. beautiful 4. barbarous 5. new	( )
14.3	*86*	*173*	*15*	*164*	45. vecino	1. vicinity 2. twentieth 3. neighbor 4. vacant 5. need	( )
46.5	*38*	*57*	*24*	*122*	46. subir	1. ascend 2. lower 3. sublet 4. substance 5. surprise	( )
78.5	*10*	*10*	*−13*	*223*	47. catorce	1. forty 2. expense 3. wildcat 4. fourteen 5. letter	( )
18.8	*72*	*148*	*10*	*184*	48. consejo	1. consecrate 2. conflict 3. accident 4. advance 5. advice	( )
31.8	*59*	*107*	*23*	*125*	49. sorprender	1. surrender 2. surprise 3. put over 4. suspend 5. hold	( )
18.5	*73*	*151*	*16*	*157*	50. rogar	1. redden 2. break 3. beg 4. answer 5. whimper	( )
22.0	*66*	*132*	*46*	*58*	51. lágrima	1. brick 2. vegetable 3. liquid 4. lamp 5. tear	( )
69.3	*21*	*24*	*−5*	*220*	52. trece	1. sad 2. third 3. three 4. thirteen 5. train	( )
37.3	*45*	*91*	*15*	*163*	53. calidad	1. expensive 2. quality 3. heat 4. furnace 5. quantity	( )
18.5	*74*	*152*	*17*	*152*	54. atravesar	1. work over 2. find out 3. surround 4. cross 5. dare	( )
45.3	*41*	*63*	*17*	*153*	55. adelantar	1. aid 2. shave 3. avarice 4. advert 5. advance	( )
44.3	*43*	*69*	*13*	*168*	56. olor	1. color 2. other 3. olive 4. odor 5. oil	( )
35.5	*52*	*94*	*7*	*194*	57. orden	1. ardent 2. garden 3. order 4. ordain 5. burn	( )

1	2	3	4	5			
17.5	*79*	*159*	*10*	*183*	58. trigo	1. trigger 2. journey 3. wheat 4. adversity 5. thrashing machine	( )
20.0	*70*	*142*	*28*	*103*	59. paraguas	1. package 2. asparagus 3. overshoes 4. place 5. umbrella	( )
47.3	*37*	*55*	*4*	*201*	60. ausente	1. although 2. awful 3. absent 4. scarcely 5. lately	( )
40.0	*49*	*107*	*43*	*64*	61. cuidado	1. guide 2. care 3. question 4. comply 5. quizz	( )
15.0	*85*	*172*	*28*	*104*	62. rato	1. rat 2. while 3. amount 4. ratify 5. rest	( )
79.3	*9*	*9*	*−5*	*219*	63. vinagre	1. wine 2. vinegar 3. sour 4. disagreement 5. vintage	( )
55.5	*31*	*38*	*0*	*213*	64. piel	1. foot 2. skin 3. then 4. coin 5. fragment	( )
17.0	*80*	*162*	*22*	*130*	65. secar	1. seek 2. acknowledge 3. sand 4. recognize 5. dry	( )
17.8	*77*	*156*	*20*	*139*	66. señalar	1. regret 2. feel 3. help 4. stamp 5. indicate	( )
16.8	*81*	*163*	*19*	*142*	67. sonrisa	1. surprise 2. smile 3. sunrise 4. dream 5. explosion	( )
72.5	*13*	*16*	*9*	*186*	68. gris	1. grease 2. grey 3. clever 4. weak 5. grizzly	( )
15.8	*84*	*169*	*6*	*196*	69. cuanto	1. many 2. more 3. how 4. as much 5. however	( )
6.8	*99*	*204*	*12*	*174*	70. propio	1. owner 2. position 3. pleasing 4. fitting 5. property	( )
44.3	*44*	*70*	*0*	*214*	71. vivo	1. voice 2. life 3. true 4. alive 5. slow	( )
28.8	*62*	*114*	*14*	*166*	72. saludable	1. salad 2. exit 3. healthful 4. greeting 5. prosperous	( )
12.5	*91*	*180*	*13*	*169*	73. recordar	1. cover up 2. receive 3. record 4. recover 5. recall	( )
11.8	*92*	*182*	*22*	*131*	74. blando	1. blanket 2. round 3. blank 4. light 5. soft	( )
14.3	*87*	*174*	*21*	*133*	75. suave	1. insipid 2. brave 3. severe 4. gentle 5. torpid	( )
61.0	*27*	*32*	*6*	*195*	76. heno	1. hay 2. deed 3. four 4. hard 5. today	( )
8.8	*97*	*195*	*10*	*182*	77. roto	1. rote 2. record 3. roast 4. round 5. broken	( )
16.3	*83*	*167*	*20*	*137*	78. sábana	1. banana 2. Saturday 3. sheet 4. savings 5. napkin	( )

1	2	3	4	5			
39.0	*50*	*86*	*11*	*177*	79. cruzar	1. cross 2. correct 3. crucify 4. believe 5. grow	( )
54.0	*32*	*39*	*8*	*188*	80. situado	1. event 2. located 3. position 4. home 5. siege	( )
14.3	*88*	*175*	*24*	*123*	81. averiguar	1. avarice 2. find out 3. address 4. inform 5. confound	( )
12.8	*90*	*179*	*41*	*72*	82. peligro	1. fight 2. pelican 3. danger 4. coiffure 5. hairy	( )
18.0	*75*	*154*	*37*	*81*	83. mentira	1. lie 2. misdeed 3. mantle 4. engine 5. mention	( )
16.8	*82*	*164*	*14*	*167*	84. afeitar	1. finish 2. slice 3. celebrate 4. shave 5. cut	( )
33.5	*56*	*98*	*26*	*119*	85. borrar	1. borrow 2. sweep 3. erase 4. wash 5. bring	( )
22.8	*63*	*129*	*26*	*120*	86. seco	1. thirst 2. dry 3. sack 4. silk 5. drought	( )
11.8	*93*	*183*	*20*	*138*	87. principio	1. sovereign 2. novice 3. principality 4. authority 5. beginning	( )
19.3	*71*	*146*	*27*	*111*	88. rodear	1. sneer 2. to hate 3. surround 4. to tear 5. revolve	( )
33.5	*57*	*99*	*3*	*203*	89. puesto	1. soon 2. harbor 3. position 4. door 5. fitting	( )
17.8	*78*	*157*	*0*	*215*	90. viudo	1. survivor 2. widower 3. old 4. widow 5. vice	( )
64.3	*24*	*28*	*−9*	*222*	91. botón	1. pump 2. button 3. baton 4. pocket 5. bottom	( )
7.3	*98*	*200*	*10*	*181*	92. alrededor	1. already 2. altar 3. away 4. hunter 5. around	( )
11.5	*95*	*185*	*7*	*193*	93. campesino	1. campus 2. peasant 3. tired 4. village 5. small camp	( )
34.0	*55*	*97*	*18*	*148*	94. deuda	1. owner 2. contract 3. doubt 4. bill 5. debt	( )
22.8	*64*	*130*	*−23*	*225*	95. tender	1. stretch 2. tender 3. delay 4. hold 5. soften	( )
11.8	*94*	*184*	*9*	*187*	96. placer	1. place 2. assist 3. star 4. plague 5. pleasure	( )
22.0	*67*	*133*	*42*	*67*	97. cereza	1. basket 2. plum 3. ceiling 4. cherry 5. cereal	( )
11.0	*96*	*188*	*4*	*200*	98. éxito	1. extra 2. exigency 3. unit 4. success 5. excitement	( )
14.0	*89*	*177*	*19*	*143*	99. calcetín	1. sock 2. cherry 3. calcimine 4. shirt 5. calcium	( )
2.0	*100*	*219*	*−1*	*218*	100. codo	1. lame 2. code 3. elbow 4. wrist 5. finger	( )

Per cent right answers ($N = 400$)	Order of difficulty within one part	Order of difficulty all three parts	Index of goodness	Order of goodness whole test (three parts)
1	2	3	4	5
75.8	1	13	59	25
52.0	9	45	70	6
45.5	18	62	69	9
49.3	12	35	67	11
62.8	4	31	55	33
75.5	2	14	36	83
53.0	8	42	67	12
60.0	5	33	50	47

PART II: SPANISH COMPREHENSION TEST

DIRECTIONS: *In each of the following statements there are five alternatives. Select the one which makes the statement most reasonable and put its number in the parentheses at the right, as in the example. You have thirty minutes for this part of the examination.*

EXAMPLE:

El gato es — 1 un libro, 2 un árbol, 3 un animal, 4 una hoja, 5 una botella. . (3)

1. El perro normal tiene — 1 una cabeza, 2 cinco pies, 3 un reloj, 4 tres orejas, 5 cincuenta años ..... ( )
2. En América los ciudadanos suelen entrar en sus casas y salir de ellas — 1 secretamente, 2 de noche, 3 por una ventana, 4 por una puerta, 5 a caballo ..... ( )
3. Supongamos que Vd. tiene una onza de cada uno de los siguientes metales; ¿cuál es de mayor valor? — 1 plata, 2 cobre, 3 acero, 4 oro, 5 zinque ..... ( )
4. Cuando uno tiene ganas de comer se dice que — 1 tiene hambre, 2 quiere volar, 3 tiene sueño, 4 hace calor, 5 desea jugar ..... ( )
5. El mes que sigue a julio es — 1 el lunes, 2 agosto, 3 ocho días, 4 un animal, 5 mayo ..... ( )
6. Los tinteros que están en uso contienen ordinariamente — 1 agua, 2 leche, 3 vino, 4 tiza, 5 tinta ..... ( )
7. La estación más fría del año se llama — 1 el verano, 2 la semana, 3 el noviembre, 4 el invierno, 5 la mañana ..... ( )
8. La mantequilla se hace — 1 del pan, 2 de la crema, 3 del agua, 4 de la sopa, 5 de la fruta ..... ( )

1	2	3	4	5	
49.0	*13*	*50*	*29*	*101*	9. Un sinónimo de "ocho días" es — 1 un mes, 2 una semana, 3 un siglo, 4 un año, 5 una vara.....( )
42.8	*24*	*76*	*58*	*27*	10. Cuando un convidado dice verdaderamente que ha comido bastante, quiere decir que — 1 tiene hombre, 2 está satisfecho, 3 tiene miedo, 4 es extranjero, 5 desea más carne.....( )
40.0	*26*	*81*	*67*	*13*	11. La silla sirve como — 1 sello, 2 libro, 3 asiento, 4 pared, 5 venta.....( )
45.0	*19*	*64*	*41*	*73*	12. Juan solamente tiene mi libro; ¿de quién es el libro que tiene Juan? — 1 el suyo, 2 el nuestro, 3 a mi, 4 el otro, 5 el mío.....( )
43.5	*23*	*74*	*53*	*35*	13. El padre de mi amigo es el propietario de una sastrería muy grande y muy buena; vaya Vd. a su oficina cuando piense mandar hacer — 1 un vestido, 2 un libro, 3 una cena, 4 un reloj, 5 un retrato.....( )
54.0	*7*	*40*	*42*	*69*	14. La pera es — 1 un animal, 2 una piedra preciosa, 3 una fruta, 4 un metal, 5 un prisonero.....( )
37.8	*30*	*88*	*59*	*26*	15. El objeto que contribuye luz y energía al mundo y que hace posible todas las formas de vida que se encuentran en nuestro hogar terrestre es — 1 la luna, 2 el día, 3 el sol, 4 la noche, 5 el crepúsculo.....( )
32.5	*34*	*103*	*53*	*97*	16. María está llorando; lágrimas amargas chorrean de sus ojos; puede suponerse que ella está — 1 contenta, 2 muy alegre, 3 indiferentemente bien, 4 riendo, 5 sufriendo mucho.....( )
45.0	*20*	*65*	*38*	*79*	17. El que fuma sin cesar es uno que usa mucho — 1 el sombrero, 2 el techo, 3 los ojos, 4 el tabaco, 5 el azúcar.....( )

1	2	3	4	5	
52.0	*10*	*46*	*51*	*42*	18. El color más brillante de la bandera americana es — 1 blanco, 2 negro, 3 amarillo, 4 moreno, 5 rojo.....( )
31.5	*36*	*109*	*43*	*65*	19. *Sepa* y *supe* son formas del verbo — 1 separar, 2 soñar, 3 sonar, 4 saber, 5 sudar.....( )
42.5	*25*	*79*	*64*	*15*	20. Por lo común el café se toma en — 1 un plato, 2 una taza, 3 el azúcar, 4 el libro, 5 un sueño.....( )
39.8	*27*	*83*	*35*	*85*	21. El chaleco es una parte — 1 del sombrero, 2 de la mano, 3 del cuerpo, 4 del vestido, 5 de la cabeza.....( )
23.8	*42*	*124*	*44*	*61*	22. La maleta se usa mucho por las personas que — 1 hacen viajes, 2 viven siempre en un lugar, 3 salen poco, 4 pasan toda la vida en el mismo hogar, 5 son campesinos.....( )
13.5	*56*	*178*	*16*	*159*	23. El señor Pereda perdió el brazo derecho en la Gran Guerra; el que le queda es — 1 el que perdió, 2 los ojos, 3 el fusil, 4 la batalla, 5 el izquierdo.....( )
16.8	*54*	*165*	*31*	*95*	24. La mayoría de los ciudadanos que viven en un país inmediatamente al sur de los EE.UU. de América son — 1 chinos, 2 ingleses, 3 brasileños, 4 mejicanos, 5 peruános.....( )
47.5	*15*	*52*	*39*	*77*	25. Muchas personas están acostumbradas a tomar el café con — 1 aceite, 2 una navaja, 3 azúcar, 4 vinagre, 5 jabón.....( )
48.5	*14*	*51*	*56*	*30*	26. Para calentar las casas en invierno se usa mucho — 1 la carabina, 2 el carbón, 3 la poesía, 4 el ruido, 5 la fiebre.....( )

1	2	3	4	5	
44.5	*21*	*67*	*63*	*18*	27. Tenga la bondad de indicar el nombre de la estación más calurosa, supuesto que se trata del clima de Madrid — 1 el lunes, 2 la semana, 3 el verano, 4 el invierno, 5 el ferrocarril.....( )
37.8	*31*	*89*	*60*	*24*	28. El azúcar es — 1 amargo, 2 dulce, 3 negro, 4 peligroso, 5 grande.....( )
20.5	*47*	*140*	*19*	*145*	29. El buque pertenece a una clase de — 1 botes, 2 pájaros, 3 ríos, 4 paraguas, 5 tranvía.....( )
23.3	*44*	*127*	*28*	*106*	30. El señor volvió a su casa y se quitó el polvo del vestido con — 1 un fósforo 2 una nube, 3 una piedra, 4 una naranja, 5 un cepillo.....( )
65.5	*3*	*27*	*45*	*60*	31. Un animal más pequeño que el perro es — 1 el león, 2 el tigre, 3 el camello, 4 el toro, 5 el gato.....( )
46.5	*17*	*58*	*51*	*43*	32. Casi todos los libros que se han publicado receientemente se han impreso con tinta — 1 rosa, 2 negra, 3 gris, 4 verde, 5 azul.....( )
17.8	*52*	*158*	*25*	*121*	33. Bajando la escalera ayer mi querido nieto cayó desde arriva y se quebró el hueso grande de la pierna derecha; — 1 nostoros noso reímos, 2 yo me puse a cantar alegremente, 3 hicimos venir al médico, 4 me burlé de él, 5 el médico dijo que era una broma bien divertida.....( )
38.0	*29*	*87*	*55*	*34*	34. Para comer la sopa el joven bien instruído se servirá de — 1 un cuchillo, 2 un tenedor, 3 una cuchara, 4 una jarra, 5 una cubeta.....( )
32.8	*33*	*102*	*51*	*44*	35. El hombre atrevido es uno que — 1 es valiente, 2 tiene miedo, 3 sabe mucho, 4 ataca a los débiles, 5 es de odíarse.....( )
24.3	*40*	*122*	*40*	*76*	36. El hielo es — 1 blando, 2 frío, 3 liquido, 4 caliente, 5 siempre inútil.....( )

1	2	3	4	5	
52.0	*11*	*47*	*61*	*23*	37. Si uno tiene sed, lo que debe hacer es — 1 tomar vinagre, 2 acostarse, 3 burlarse de sus amigos, 4 tomar agua, 5 correr a más no poder.....( )
59.8	*6*	*34*	*37*	*82*	38. ¿Cuántos continentes hay en el mundo? — 1 cuarenta, 2 cincuenta, 3 trece, 4 diecisiete, 5 cinco.....( )
44.5	*22*	*68*	*48*	*54*	39. Juanito tiene doce años, es decir, tres años más de los que lleva su hermana Rosita; ¿qué edad tiene Rosita? — 1 ocho años, 2 nueve meses, 3 trece años, 4 nueve siglos, 5 nueve años.....( )
39.5	*28*	*84*	*44*	*62*	40. Cuando un caballero bien educado se descubre cortesmente en presencia de una señora que ha encontrado en el paseo, qué se quita? — 1 un anillo, 2 los guantes, 3 la corbata, 4 los anteojos, 5 el sombrero.....( )
22.0	*46*	*134*	*35*	*86*	41. Juan les dijo que no tenía los libros; a quiénes dijo esto Juan? — 1 a mí, 2 a él, 3 a ellos, 4 a si mismo, 5 a los libros.....( )
30.5	*38*	*111*	*42*	*70*	42. Mi amigo, Juan Valera, firmó con su nombre la carta que Carlos había escrito; Juan escribió en la carta — 1 Carlos, 2 mi amigo, 3 Juan Valera, 4 la carta, 5 había.....( )
20.0	*48*	*143*	*33*	*92*	43. En cierta, estación del año la gente suele llevar ropa muy pesada; porqué lleva ropa muy gruesa? — 1 es malo, 2 hace calor, 3 la cortesía lo exige, 4 hace frío, 5 llueve mucho.....( )
23.5	*43*	*126*	*26*	*114*	44. Mandaron hacer el trabajo a Juan, pero él se encontraba indispuesto y lo hice yo; y trabajé en lugar de — 1 hacer, 2 Juan, 3 indispuesto, 4 el trabajo, 5 encontraba.....( )

1	2	3	4	5	
47.5	*16*	*53*	*51*	*45*	45. El horario de mi reloj está entre las nueve y las diez y el minutero indica las seis; ¿qué hora es según dicho reloj? — 1 las seis, 3 las nueve y media, 4 mi reloj, 5 según ..... ( )
33.3	*32*	*100*	*53*	*37*	46. La batalla de Lepanto fué sangrienta; muchos soldados sufrieron heridas graves, y otros muchos perdieron — 1 un lápiz, 2 la vida, 3 un libro, 4 una copa, 5 un navío ..... ( )
17.3	*53*	*160*	*26*	*115*	47. Los niños, Juan, Carlos y Miguel, se pusieron las gorras, cada cual su propria gorra; ¿qué gorra se puso Juan? — 1 el suyo, 2 el otro, 3 la nuestra, 4 la suya, 5 el propio ..... ( )
31.0	*37*	*110*	*49*	*50*	48. El cuarto es tan ancho como largo, es decir, tiene quince metros de ancho; cuántos metros de longitud tiene la pieza — 1 quince, 2 cinco, 3 tres, 4 ancoh, 5 cuarto ..... ( )
32.0	*35*	*105*	*65*	*14*	49. Ayer fué el día primero del primer mes del año; a cuantos estamos hoy? — 1 a últimos de Diciembre, 2 dos de enero, 3 segundo de mayo, 4 el primero de diciembre, 5 dos de febrero ..... ( )
28.8	*39*	*115*	*31*	*60*	50. Si hoy es el veintiocho de febrero (1925); ¿qué fecha será mañana? — 1 el veintinueve de febrero, 2 el primero de marzo, 3 el primero de abril, 4 el último del mes, 5 el día natal ..... ( )
23.3	*45*	*128*	*42*	*71*	51. El que tiene deudas honestas y dinero debe — 1 olvidarlas, 2 pagarlas, 3 matar a sus acreedores, 4 tardar demasiado, 5 esconderse ..... ( )
16.0	*55*	*168*	*20*	*141*	52. En un naufragio los que corren menos peligro de perder la vida son los que — 1 saben nadar, 2 están enfermos, 3 están dormidos, 4 tienen miedo, 5 están locos de miedo ..... ( )

1	2	3	4	5	
24.0	*41*	*123*	*31*	*97*	53. Paseándose por los campos un día Juan notó que se alzaba rápidamente una nube muy negra y muy grande; a los pocos momentos empezaron a caer unas grandes gotas de — 1 polvo, 2 agua, 3 viento, 4 vino, 5 miel.....( )
10.3	*58*	*190*	*22*	*132*	54. Los pájaros vuelan por el aire por medio de sus — 1 pies, 2 alas, 3 voces, 4 canciones, 5 patitas.....( )
9.3	*60*	*194*	*16*	*160*	55. Hay unos cajoncitos que los fumadores abren cuando quieren encender un cigarro; estos cajoncitos contienen — 1 tabaco, 2 cerillas ó fósforos, 3 tinta, 4 dinero, 5 velas.....( )
11.3	*57*	*186*	*11*	*178*	56. Cuando uno se ha puesto el sobretodo puede decirse que lleva — 1 un paquete, 2 dinero, 3 abrigo, 4 dos cuellos, 5 tres guantes.....( )
18.8	*50*	*149*	*27*	*112*	57. El maestro estaba escribiendo con una pluma de — 1 papel, 2 madera, 3 pimienta, 4 paja, 5 oro.....( )
9.5	*59*	*193*	*11*	*179*	58. Las personas que siempre ayudan a los pobres y que no pierden ocasión de proteger a los débiles — 1 deben ser alabadas, 2 son culpables, 3 deben perder la vida, 4 son muy numerosas, 5 son criados ineptos.....( )
18.8	*51*	*150*	*21*	*134*	59. Si un hombre está nadando está necesariamente — 1 en el mar, 2 en un hospital, 3 en el agua, 4 enfermo, 5 en un tren.....( )
19.8	*49*	*144*	*12*	*175*	60. Un animal que ladra a veces, es muy útil en las haciendas — 1 el gato, 2 el gallo, 3 el perro, 4 el caballo, 5 el burro.....( )

PART III: SPANISH COMPLETION TEST

DIRECTIONS: *Each of the English sentences below is followed by a translation which is correct as far as it goes, but which is incomplete. Each translation can be completed correctly by inserting one, two or three words in each blank* (——). *When you have decided what is necessary to complete the translation, write it on the line at the right of the translation. WRITE LEGIBLY. DO NOT SPEND TOO MUCH TIME on any one sentence. Do all the easy ones first, then go back to the harder ones if you have time. You will have thirty minutes for this part of the examination. Write carefully and clearly, putting your answers on the lines at the right as in the examples.*

EXAMPLES:

1. I am a man. (——) hombre ..... Soy
2. He is here. El está (——) ..... aquí

Per cent right answers (N = 400)	*Order of difficulty within one part*	*Order of difficulty all three parts*	*Index of goodness*	*Order of goodness whole test (three parts)*	
1	2	3	4	5	
45.8	*4*	*61*	*70*	*7*	1. I am 12 years old. (——) doce años. .....
37.5	*9*	*90*	*63*	*19*	2. How old are you? (——) tiene Vd? .....
37.3	*10*	*92*	*80*	*3*	3. I wrote to my parents yesterday. Ayer escribí (——) padres. .....
32.5	*12*	*104*	*70*	*8*	4. Do you want to go to the theatre? ¿Quiere (——) al teatro? .....
44.3	*5*	*71*	*49*	*51*	5. Don't mention it. No hay (——) .....
25.3	*20*	*121*	*64*	*16*	6. I was studying and my brother was reading. Yo estudiaba y mi hermano (——) .....
46.3	*3*	*56*	*73*	*5*	7. Put it on the table. Póngalo (——) mesa .....
21.8	*22*	*135*	*44*	*63*	8. They are very warm. Ellas (——) mucho calor. .....

1	2	3	4	5	
39.3	*8*	*85*	*47*	*57*	9. Where are they? (——) están?
32.0	*13*	*106*	*81*	*1*	10. It will be good weather tomorrow. Hará buen (——) mañana
74.0	*1*	*15*	*24*	*124*	11. The book is old. El libro (——) viejo
18.5	*28*	*153*	*50*	*48*	12. I read this book yesterday. (——) este libro ayer
27.3	*16*	*117*	*62*	*22*	13. All the men are here. Todos los hombres (——)
26.0	*17*	*118*	*63*	*20*	14. What is your name? (——) llama Vd?
42.8	*6*	*77*	*81*	*2*	15. It is 5 o'clock P.M. (——) cinco de la tarde
30.5	*14*	*112*	*53*	*38*	16. Fanny is from Colombia. Paquita (——) Colombia
53.5	*2*	*41*	*51*	*46*	17. This is a wooden chair. Esta silla es (——)
23.8	*21*	*125*	*43*	*66*	18. How much money have you? (——) dinero tiene Vd?
20.8	*24*	*138*	*63*	*21*	19. How do you like this bread? Cómo (——) este pan?
19.5	*26*	*145*	*39*	*78*	20. I shall leave tomorrow night. Partiré mañana (——) noche
27.8	*15*	*116*	*69*	*10*	21. They are mine. Son (——)

1	2	3	4	5	
15.8	*31*	*170*	*53*	*39*	22. He gives it to you. (——) da a Vd.
37.0	*11*	*93*	*77*	*4*	23. They are ours. Son (——)
6.5	*46*	*205*	*17*	*155*	24. Do you feel better now? Se siente Vd. (——)?
25.5	*19*	*120*	*57*	*29*	25. John is taller than I. Juan es (——) que yo.
26.0	*18*	*119*	*64*	*17*	26. He lived thirty years. (——) treinta años.
3.0	*58*	*217*	*8*	*190*	27. I have just read it. (——) leerlo.
19.3	*27*	*147*	*49*	*52*	28. Give me a little bread. Déme Vd. (——) de pan.
14.3	*33*	*176*	*34*	*91*	29. Where are they going? (——) van?
11.3	*35*	*187*	*28*	*107*	30. They were tired. (——) cansados
8.3	*40*	*197*	*28*	*108*	31. The children were sleepy. Los niños (——)
5.8	*48*	*207*	*11*	*180*	32. He comes very often. Viene muy (——)
5.5	*51*	*199*	*19*	*146*	33. He put on his hat. Él (——) el sombrero
2.0	*60*	*220*	*8*	*191*	34. He began to run. (——) correr

1	2	3	4	5	
15.5	*32*	*171*	*41*	*74*	35. He fell near the trunk. Se cayó (——) baúl
6.3	*47*	*206*	*23*	*129*	36. Fruits are sold in the market. Las frutas (——) en el mercado
21.0	*23*	*136*	*48*	*55*	37. John is her oldest brother. Juan es su hermano (——)
12.3	*34*	*181*	*36*	*84*	38. I have more than five pesos. Tengo (——) cinco pesos
5.8	*49*	*208*	*16*	*161*	39. This man, although poor, is honorable. Este hombre, (——) pobre, es honrado
9.8	*38*	*192*	*29*	*102*	40. They were under the table. Estaban (——) mesa
20.3	*25*	*141*	*53*	*40*	41. It was hidden behind the bed. Estaba escondido (——) cama
1.5	*61*	*221*	*2*	*209*	42. The man went crazy. El hombre (——) loco
5.8	*50*	*209*	*21*	*135*	43. He had to leave early. (——) partir temprano
7.5	*41*	*198*	*27*	*113*	44. Don't tell it to me. No me (——)
3.8	*56*	*215*	*13*	*171*	45. After they had fought for ten years. (——) hubieron combatido diez años
10.8	*36*	*206*	*35*	*87*	46. Whom do they see? (——) ven?
17.3	*29*	*176*	*26*	*116*	47. In one hour he will be far from here. En una hora estará (——) aquí

1	2	3	4	5	
3.8	*57*	*216*	*13*	*172*	48. I wish to take off my gloves. Quiero (——) los guantes
5.0	*53*	*212*	*16*	*162*	49. I have seen nobody. No he visto (——)
5.3	*52*	*211*	*18*	*150*	50. I don't know who it is. No (——)
7.5	*42*	*199*	*26*	*117*	51. I met my father and he has told me all. Encontré a mi padre y (——) dicho todo
4.8	*55*	*218*	*15*	*165*	52. They were looking toward the village. Miraban (——) la aldea
16.5	*30*	*166*	*46*	*59*	53. Which is your pen? That one. ¿Cuál es su pluma? (——)
7.3	*43*	*209*	*28*	*109*	54. I don't know whether my teacher will come or not. No sé (——) mi maestro o no
2.8	*59*	*218*	*10*	*185*	55. Not being able to find him, he returned without him. No (——) encontrarle, volvió sin él
4.8	*54*	*213*	*19*	*147*	56. Since you went away. (——) Vd. se fué
7.0	*45*	*203*	*26*	*118*	57. To earn money one must work. (——) trabajar para ganar dinero
.8	*64*	*224*	*3*	*207*	58. I do not believe in it. No (——)
10.3	*37*	*191*	*32*	*93*	59. Whose horse is this? ¿(——) es este caballo?
8.8	*39*	*196*	*30*	*97*	60. I said, "A whole day." Dije, "Un día (——)"

1	2	3	4	5	
7.3	*44*	*202*	*18*	*151*	61. The dog is outside. El perro está (——) .....
40.0	*7*	*82*	*58*	*28*	62. There is John. Allá (——) Juan .....
.5	*65*	*225*	*2*	*210*	63. Let him come if he wants to. (——) venir si quiree .....
1.5	*62*	*222*	*5*	*199*	64. I have a water tumbler. Tengo un vaso (——) agua. .....
1.3	*63*	*223*	*3*	*208*	65. How high is the house? ¿Qué (——) la casa? .....

The first column of figures at the left of the test-questions shows the per cent of students that answered each question correctly. These per cents are based upon the responses of 400 students, 100 from each of the four normal classes. They constitute indices of the absolute difficulties of the items, and may be accepted as fairly reliable, since we learned above that the reliability of difficulty indices based upon only two hundred cases is about 0.92. Thus we learn (column 1, page 49) that *paysan* is understood by 78.8% of the 400 students whose responses were studied in detail, to mean *peasant*, and not *paint*, *country*, *plow*, or *safety*. The second French vocabulary item, *appeler*, was answered correctly by 38.8% of the students.

The second column gives the order of difficulty of the items within each Part. Thus we learn that *paysan* is the fifth easiest item in the 100-item vocabulary test; that *appeler* is the fiftieth, *écouter* the fourth, *courir* the thirty-eighth, and that *argent* is the easiest of the 100 vocabulary items. Similarly we learn, from column 2, that *La capitale de la France . . .* is the easiest item in Part II, and *donnez-moi* the easiest item in Part III of the French test. The most difficult vocabulary item is No. 93, *renoncer;* only 3% of the students answered it correctly. These few answered correctly in spite of the fact that the word *abjure* was printed as *adjure.* The most difficult item in Part II is No. 57, and in Part III No. 57, also.

The third column shows the order of difficulty of the items when all three Parts are merged. Thus *paysan* is the eighth easiest item in the 220 questions that make up the French test; *La capitale de la France . . .*, in Part II, is the easiest of all the 220 questions in the French test. The most difficult of the 220 hurdles of the French test is No. 57 in Part III.

The reader may secure some amusement, and perhaps profit, by indulging in a little guessing game as to the relative difficulties of any two contiguous items in these two tests. He will have a concrete illustration of the difficulty of constructing examinations "by guess," as old-type examinations have necessarily been constructed in the past.

Columns 4 and 5 will be explained in connection with the discussion of "good" and "bad" questions, in the following paragraphs.

Distribution of Henmon and Wood frequencies of words in French vocabulary test. — Columns 6 and 7 for Part I of the French test were put in as an additional check on the difficulty and validity of our vocabulary test items. Column 6 shows the number of times each word occurred in 400,000 words of running discourse, according to the Henmon *French Word Book* (q.v.).[1] Column 7 shows the number of textbooks[2] in which each word appeared. Table 10 summarizes the facts of columns

[1] University of Wisconsin Research Bulletin No. 3, Sept., 1924. See above, pp. 49–54.

[2] Wood, *A Comparative Study of the Vocabularies of Sixteen French Textbooks*, The Modern Language Journal, February, 1927.

6 and 7, and shows that although every word in our French vocabulary test occurs in four or more standard and widely used French textbooks, fifteen of them do not occur in the Henmon list. This means that 85 of the 100 words in Part I occur five times or oftener in 400,000 words of running discourse; and that 15 occur less than five times.

TABLE 10

DISTRIBUTION OF FRENCH WORDS IN PART I OF THE JUNIOR HIGH SCHOOL FRENCH TEST ACCORDING TO HENMON FREQUENCIES AND ACCORDING TO NUMBER OF FRENCH TEXTBOOKS IN WHICH EACH WORD OCCURS.

HENMON FREQUENCY: IN 400,000 WORDS RUNNING DISCOURSE	NUMBER OF PART I FRENCH WORDS	WOOD FREQUENCY: NUMBER OF FRENCH TEXTBOOKS	NUMBER OF PART I FRENCH WORDS
0	15	4	10
5–9	10	5	11
10–19	17	6	11
20–29	11	7	10
30–39	6	8	8
40–49	2	9	
50–59	7	10	1
60–69	3	11	1
70–79	2	12	9
80–89	4	13	9
90–99	1	14	10
100–119	4	15	9
120–139	3	16	11
140–159	2		
160–179	3		
180–199			
200–299	2		
300–399	3		
400–499	2		
500 +	3		

Intercorrelations of Henmon and Wood frequencies and empirical difficulties of words in French vocabulary test. — The correlation of the empirical order of difficulty (column 2, pp. 49–54) of these 100 French words with the Henmon frequencies of column 6 is 0.278; and with the Wood frequencies of column 7 is 0.336. The correlation between columns 6 and 7 is 0.832. These correlations have considerable significance for the content and organization of the curriculum as well as for the theory and practice of modern language testing. It is clear that under present conditions students learn many words which occur only once or twice in 400,000 words of discourse better than they learn other words which will be met with much oftener in general reading.

VALIDITY OF INDIVIDUAL QUESTIONS

Definition of a "good" question. — On the basis of the responses of the 400 cases used in establishing the order of difficulty discussed above,

each of the 220 or 225 items in the French and Spanish tests has been checked up for validity against two external criteria. The first criterion was the class-placement, 8A, 8B, 9A, or 9B, of the students; the second was the total score on the French or Spanish test. The 400 cases were divided into four groups of 100 each, Group I being composed of the 100 who secured the lowest scores, regardless of what class they were in, and Group II the next 100 lowest scores, . . . and Group IV of those who secured the highest 100 scores in the sample of 400 cases. A question is considered valid or "good" if it consistently differentiates between groups of students which are known to be arranged in order of merit, that is, if it gives progressively higher percentages of correct answers to the four groups just described, from 8A to 9B, or from I to IV, in order. A question is considered "bad" if it does not favor the better groups, and is considered very bad if it actually gives more credits to the 8A or Group I students than it gives to the 9B or Group IV students.

Possible causes of "inverted" questions. — Those who have not made exact empirical studies of examination results may be inclined to doubt that a reasonable question could be constructed which would give higher scores to poor students than to good students; but such questions occur frequently in all types of examinations. The causes of these "inversions" are complex, and vary from question to question. We shall in following pages make some suggestions as to the probable causes of the few inversions which we have found in these tests; the only generalization that seems possible is that the inversions may be genuine or false: genuine, if the advanced students have really forgotten something which they learned very well in an earlier semester; false, if the low scores of the advanced students are due to some subtle ambiguity in the question itself, in the directions which have been given to the students, or in some prepotent counter-suggestion in the setting of the question which takes effect with the bright students but passes unnoticed by the dull ones. There can be no doubt that advanced students do forget some things they have learned well in early semesters, because of inadequate articulation of lower with higher class work, and because of the efforts of teachers to follow too mechanically an artificially segmented syllabus.

"Inversions" should be detected and eliminated. — Whatever the cause of "inverted" questions may be, "inversions" should be detected and eliminated from all examinations the results of which are likely to affect the education of any number of students. It is due to the presence of such questions in old-type examinations, among other unfortunate features, that old-type examinations afford such unsatisfactory reliability and validity coefficients. The purpose of an examination is to separate the sheep from the goats; and this is the purpose of each question in an examination. A question which gives equal credits to all students, low

and high, does not serve the purpose for which the examination was made and administered; but it does no particular harm, aside from wasting the time and energy of both students and teachers. If, however, a question actually gives higher scores to poor students than to good students, its influence runs counter to the purpose of the examination; it pulls the sheep downward, and pushes the goats upward. Inversions cannot be detected without actual trial, and no examination should be accepted for use on a city- or state-wide basis which has not stood the test of experimental verification.

TABLE 11

Validities of Individual Questions in French and in Spanish Tests

SHOWING THE NUMBERS OF ITEMS IN EACH PART OF EACH TEST OF EACH DEGREE OF "GOODNESS" OR "BADNESS," AS DEFINED IN THE TEXT. THE FIRST COLUMN, HEADED IV–I, SHOWS THE INDICES OF "GOODNESS" IN TERMS OF THE DIFFERENCE BETWEEN "PER CENT OF CORRECT RESPONSES BY GROUP IV STUDENTS" AND "PER CENT OF CORRECT RESPONSES BY GROUP I STUDENTS."

	French Test			Spanish Test		
IV%–I%	Part I	Part II	Part III	Part I	Part II	Part III
80–90	3	4	1			3
70–79.9	8	11	6		1	4
60–69.9	17	15	7		9	7
50–59.9	16	11	12	3	14	7
40–49.9	5	10	9	9	11	8
30–39.9	11	3	8	6	10	6
20–29.9	14	3	4	22	9	11
10–19.9	15	2	8	27	6	12
5–9.9	2		1	11		3
0–4.9	4	1	4	12		4
−1 to −4.9	3			3		
−5 to −9.9	2			4		
−10 to −19.9				2		
−20 to −29.9				1		

As a rough measure of the goodness or badness of individual questions we have compared the percentages of 9B (or Group IV) students with the percentages of 8A (or Group I) students that answered each question correctly. If a larger proportion of the 9B students answer a given question correctly than of 8A students, the question is good; and its degree of goodness is roughly indicated by the amount of the difference. Column 4 shows for a given question in each test the difference obtained by subtracting the per cent of Group I students answering it correctly from the per cent of Group IV students answering it correctly. (See above, pp. 49–85. The 9B–8A differences were found to be so closely parallel to the IV–I differences that only the latter need be presented.) If a figure in column 4 is positive, the question to which it applies is a good one,

for it means that good students answered it correctly more often than poor students; if a figure is preceded by a minus sign (−), the question is a bad one, because poor students answered it correctly more often than good ones. Indifferent questions are those which give only small positive differences in favor of the advanced group of students. A summary of the facts of column 4 is presented in Table 11.

Proportions of "good" and "bad" questions in the tests. — From the first entry in the column headed French Part I, in Table 11, we learn that there were three French vocabulary items which were answered correctly by an excess of Group IV over Group I students of between 80 and 90 per cent. The second entry shows that there were eight vocabulary items for which the excess of correct answers in Group IV over correct answers in Group I was between 70 and 80 per cent. In the whole French test there are only 17 items for which the excess of Group IV correct responses over Group I correct responses is less than 10%; in other words, 203, or over 92% of the 220 items in the French test, wield effective differentiating power, and only 5, or 2.3%, of the 220 items differentiate in the wrong direction, and these 5 items are only slightly inverted.

Practically the same proportions hold for the Spanish test in so far as Parts II and III are concerned; the Spanish vocabulary test is less fortunately constructed. No Spanish vocabulary item yields a difference in favor of Group IV above 60%, and only 67 of the vocabulary items wield significant differentiating influence, and 10 wield at least a slight influence in the wrong direction. On the whole, the difference between the French and Spanish tests here displayed is not very great; Table 11 indicates that the items of both tests have a very high "batting average"; but it is interesting to note that Table 11 favors the French test to about the same extent as the reliability coefficients of Tables 6 and 7 do. This is a striking example of internal or analytical evidence confirming total correlation results, and illustrates the importance of and the need for meticulous care in validating each and every individual item in a test. Only by such care can we construct tests such that each item and each moment of examination time will contribute to the purpose of the examination in terms of real and effective differentiation between the various levels of achievement from lowest to highest.

Analysis of some "bad" questions. — Column 5 shows the order of goodness of all the items in each test. The item in each test which differentiates between the poorest and the best students most adequately is numbered 1 in column 5; and the item in each test which most tended to give advanced students low scores and poor students high scores is numbered 220 in the French and 225 in the Spanish test. (See pp. 49–85.)

The worst of all the items in the two tests is Spanish vocabulary item No. 95, *tender, to stretch.* The excess of poor students over good students

who answered this item correctly is 23%. The reader needs only to look at this item in order to convince himself that it is a very faulty question: but it is very doubtful whether any Spanish scholar would have picked it as a bad one in a casual, or even in a critical, inspection of the test. This item, and several others like it, escaped the vigilance of several scholars who were kind enough to review our tests critically. At least two of the alternative English words contain illegitimate seductive qualities: *tender* is spelled exactly like the Spanish word, *tender;* and *hold* is the equivalent of the Spanish word *tener,* for which *tender* might be easily mistaken by the good student who is working at top speed.

The next few worst items in the Spanish test, vocabulary items 24, 47, 91, 19, 52, 63, and 100, all contain unnecessary and questionable seductive elements in the English alternatives. The worst items in the French test are in Part I, Nos. 40, 94, 85, 91, and 68. The faults of the French items are less obvious than those of the Spanish vocabulary items; the reader may verify this judgment by comparing the French and Spanish items in detail.

Graphs of "good" and "bad" questions. — The method described above for detecting good and bad test elements is only approximately adequate; we have used it and presented the results of such use only as a rough means of showing concretely the need for studying individual items carefully in constructing a test which shall be valid and reliable. In order to be really good, a test question must not only show a difference of correct responses in favor of the highest quarter in comparison with the lowest quarter of the students, but must give progressively greater percentages of correct answers as we ascend the scale of achievement, and the increments in correct responses should be fairly regular and nearly proportional. In Chart 14 are shown graphs of the increments in correct responses from the first, second, third, and highest quarters, for the best question in each test.

The first increment curve, labelled (a), is for French vocabulary item No. 60; it is about the most ideal question that it has been my fortune to deal with. About 7% of the lowest quarter of the students answer it correctly; the per cents then go up to 32% for the second quarter, 66% for the third, and 93% for the highest quarter. Curve (e) shows that item No. 4 in Part II of the Spanish test is also very nearly ideal. Curves (c) and (f) show practically no increment of correct responses from the first to the second quarter, or from the 8A to the 8B semester. This is quite characteristic of many Part III items in both tests; students in junior high school do not apparently begin to learn to write in a foreign language until after the second semester.

Unfortunately, the increment curves of Chart 14 are not quite typical for our tests; but, as may be surmised from Table 11, the typical curves

for our test items are much nearer to these than they are to those shown in Chart 15. In the latter chart we show the increment, or rather *decrement*, curves for the worst item in each Part of each test.

The first curve in Chart 15 shows that French vocabulary item No. 94 is answered correctly by about 34% of the lowest quarter, 42% of the second quarter, 40% of the third quarter, and by only about 26% of the highest

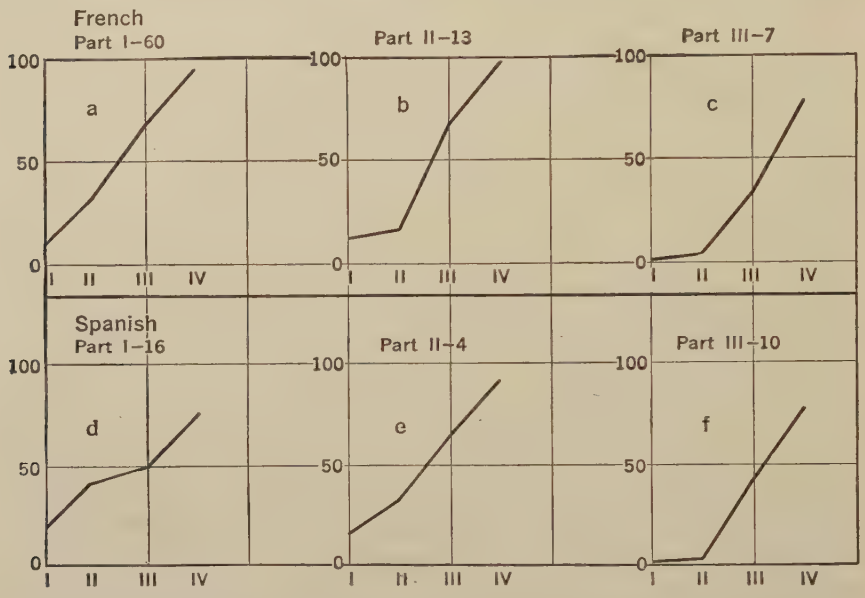


Chart 14. — Showing for the best question in each Part of each test, the percentage of correct responses from each quarter of an unselected sample of 400 students. The sample includes 100 students from each of the four normal semester groups. Group I is made up of the 100 students in the sample who secured the 100 lowest scores; Group II of students making the next 100 lowest scores; Group IV includes the students in the sample of 400 who made the highest 100 scores. The increment curves when the abscissae represent the 8A, 8B, 9A, and 9B classes practically coincide with the curves as shown here. Curves (a), (b), and (c) relate to French test items; (d), (e), and (f) to Spanish test items.

quarter! The corresponding per cents for 8A, 8B, 9A, and 9B classes are 39, 40, 59, and 23, respectively. There is no apparent trick element in this item, so that it is quite possible that students learned the word *démontrer* in the first semester, retained it through the third semester, and forgot it, to a certain extent, through disuse during the fourth semester. This tendency is still more pronounced in Spanish vocabulary items Nos. 91 and 95, although in these items there are counter-suggestions to the students which may account for their being inverted. However, these graphs and remarks are included here merely as suggestions which may

help in disseminating a more adequate appreciation of both the importance and the very real difficulties of constructing valid and reliable modern language tests. Studies of the consistency and economy and

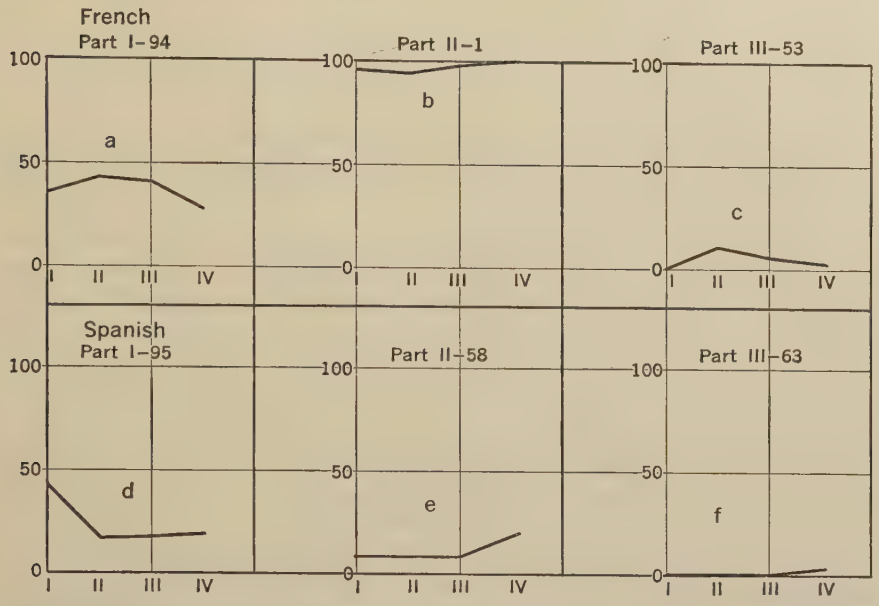


Chart 15. — Graphs illustrating the lack of validity of some of the least effective questions in the new-type French and Spanish tests.

articulation of modern language courses can be put on a sound basis only when it is possible to administer objective, reliable, and comparable tests to the *same* students at various stages of their progress in modern language work.

V

SUMMARY AND RECOMMENDATIONS

Importance and difficulty of constructing good tests. — In immediately preceding paragraphs we have compared the percentages of correct responses of several groups of different students, and not of the same students at different times. Real knowledge about the way in which our students learn, or fail to learn, a modern language as presented in high school and college courses will be achieved only when we have traced the specific modern language achievement of the same individual students through several years of effort by means of comprehensive and comparable tests frequently administered. This has not been possible thus far because of the subjectivity and unreliability of the traditional forms of examinations. A large part of the weakness of the old-type examinations is inherent in their form; but some of it at least is due to the casualness with which they are constructed. The usual run of old-type modern language examinations are constructed by a single teacher, or by a committee of two or three or four teachers, in one or two hours. The French and the Spanish tests with which we are concerned in this report are, on the one hand, a part of the results of four years of continuous experimenting with various forms of questions and of analytical work with French and Spanish textbooks; and, on the other, are the specific result of two months of work during which preliminary lists of questions were administered to several thousand junior and senior high school students. The prompt casualness with which old-type examinations are made denotes either a lack of appreciation of the importance of good measurements, or of the real difficulties of making passably reliable examinations, or both.

The essential purpose of examinations. — One of the stumbling blocks in the way of progress in examination making has been the widely accepted idea that students should be examined or tested only upon those particular and specific parts of the subject matter upon which the teacher had instructed them, or *thought* he had instructed them. It has been thought that any other kind of examination would be *unfair*, and that theoretically it should be possible for any student who had done his work acceptably to make a mark of 100% on every examination given to him. There is, in fact, no objection to the teacher's amiable desire to know how much of his or her particular offerings the students have absorbed; but the administration of the school, the student, the parents, and society

at large have a right to know how well a student at the end of a given course can, e.g., read French prose, regardless of whether he learned it from the teacher of that course, or in spite of the teacher. In other words, the interests of all concerned demand examinations which measure relative achievement in the subject matter, and not relative achievement in the particular *compote* which students have "had" from particular teachers. It is not a question of the fairness of the examination to the student, but of getting at *educationally significant truth and expressing that truth in understandable and usable terms.* Nothing could be more unfair to students than giving marks which are almost certain to be misinterpreted and misused, which is what happens very often when marks are assigned on the basis of individualistic or local standards.

Criteria of good examinations. — The fundamentals of good examinations are validity, reliability, comparability and administrative feasibility. These have been constantly kept in mind at every step in the construction of our tests. We have tried to make our tests *valid* by basing our questions on materials which word-counts and analyses of textbooks and of syllabi have shown to be important, if not in fact indispensable, common essentials; and by making our sampling of the common essentials adequate not only as to extent of materials, but also as to variety and depth of learning-units tapped. We have tried to make our tests *reliable* by insisting on large numbers of relatively independent questions, the answers to which involve few irrelevant activities on the part of the student, and which can be scored objectively, accurately, and expeditiously. We have tried to make our tests yield *comparable* measurements by making them *valid* and *reliable* tests adapted to the *whole range of achievement* in the first three or four years of modern language work, and by constructing several equivalent Forms at the outset in such a way that additional Forms equivalent in difficulty and variability may be made as needed. And finally, we have tried to make our tests administratively feasible by putting the questions in such a way that the tests are largely self-administered, and by arranging spaces for the students' answers such that the scoring is not only objective, but accurate and expeditious. The real cost of these tests when administered to large groups is less than half as much as that of old-type tests, which are not half as reliable and valid.

Although these objective tests require only ninety minutes to administer, they include 220 or 225 independently scored items. In the French test the vocabulary sampling alone includes 538 different root words. Old-type 90-minute tests ordinarily include a vocabulary sampling of less than 200 different root words. Disregarding duplicates, we have 100 words in Part I as a test of passive vocabulary; 477 in Part II, the reading-comprehension test of 60 carefully graded items; and 80 in Part III as an *active* vocabulary test in highly varied contexts. Part III is primarily a

grammar test of significant and varied content; but in addition to being this and an active vocabulary test, it is a fairly good test of idioms, order of words, agreements, hyphens, elisions, spelling, use of apostrophes, and other marks, etc., which are usually not covered in objective tests, and only imperfectly in old-type examinations. It is this combination of features, in brief, which makes these tests yield comparable and valid measures two or three times as reliable and only half as costly as old-type examinations of two or three hours' length yield.

Some limitations of the new-type tests. — These tests do not pretend to measure the oral and aural skills or cultural content. They are designed to measure only knowledge of the written language, and measure other things only in so far as they are implied by, or are dependent upon knowledge of vocabulary, reading and grammar. In any case, measurements of oral and aural skills and cultural content should be kept separate from one another and from measures of knowledge of the written language. They are three relatively independent functions; it is obvious that students often possess one of them without the other two, and failure or success in one should never be confused with, nor reported as, failure or success in another.

Moreover, paper-and-pencil tests of oral and aural skills do not seem practicable at the present time. The best way to measure these is by means of conversations with students, one at a time, using carefully prepared sets of questions and conversational materials. Such measures, however, would be subjective, would be mixtures of many relevant and irrelevant qualities in both student and teacher, and would not be comparable from place to place and year to year. These facts are not cited as arguments against the use of such tests, but as arguments in favor of keeping the marks from such tests separate from the marks derived from reliable and objective tests. To mix such marks by *averaging* would be about as illuminating as to average reliable measures of *height* with guessed *weights* of students in order to secure a total physical growth index.

Some objections considered. — There are few teachers who do not admit that the objective tests are better as *measuring* devices, but some teachers fear that the objective tests are pedagogically unsound, and that they will tend to mechanize teaching and produce what is called "dead uniformity." Specifically, it is feared by some that the objective passive vocabulary tests will cause students, aided and abetted by their teachers, to "memorize *mere* lists of words." Thus it is feared that students might make a serious breach in future objective tests by the simple and effective device of *memorizing* the small matter of the two or three thousand most frequently used words in each language! The reader can judge for himself whether such a result would be unpedagogical and calamitous, or an unhoped-for blessing. It may well be that the objective vocabulary test

may produce such a miracle; but the proponents of new-type tests have never been optimistic enough to hope that this charge against their tests would turn out to be true.

It is also difficult to see how these tests would tend to mechanize teaching. Indeed, it is a question whether the great weight put upon translation in the old-type examinations does or does not make a boomerang of this charge against the new-type tests. The translation work in modern language classes, after many years of trial, is certainly not above suspicion; many eminent authorities have found fault with it, and recent investigations tend to confirm their views.[1] On the other hand, there is no evidence against the objective tests: all the charges made against them seem to be bold prophecies spun out of conservative fears. There are not even any *a priori* grounds for imagining that the kind of objective tests here proposed will exercise unpedagogical influence. Good teachers will use the best available pedagogical devices, regardless of the merits of such devices for testing purposes; and so examiners should use the best available measuring devices, regardless of what their merits as teaching devices may be. Lack of measuring value does not destroy the pedagogical utility of a device; nor does lack of pedagogical value in a good measuring device destroy its value for examination purposes. But we do not need to rely on this distinction. On the contrary, it seems almost certain that the new-type tests will be as great a boon to teachers as to examiners. If new-type tests of the kind here proposed were a regular part of the examinations used in our schools, both teachers and students would know that the results of their efforts would be accurately and fairly measured, without the possibility of personal bias being exercised against them or invoked in their behalf; and both teachers and students would know that the whole test could not be passed with vocabulary alone, unsupported by knowledge of grammar, idioms, reading ability, etc. It is obvious that the use of objective tests would not necessarily preclude the judicious use of other types of tests for the measurement of modern language factors which are not taken care of by new-type forms of questions.

It is also charged that the objective tests do not measure the spiritual and cultural gains of the students from modern foreign language study. In so far as these gains are related to or dependent upon knowledge of the language itself, the objective tests measure them better than the old-type examinations do, for the reason that the former give fuller and more accurate measures of language-proficiency than the latter do. In so far as these gains are not related to or dependent upon knowledge of the modern language itself, neither type of test should be influenced by such gains. This does not mean that teachers should not judge the moral and

[1] The researches of Briggs and Miller, Woodring, Leonard, and others, are all in essential agreement in indicating the deleterious effects of translation exercises on English composition.

spiritual characters of their students, if they feel competent to do so; it does mean that such judgments should not be called French, or Spanish, or German, etc. That the objective tests do not mix moral appraisals with measures of defined achievement is one of their greatest merits; it is certainly not a defect to avoid confusion and obscurity in our school grades.

"Dead uniformity" is another specter which some critics pretend to see in the wake of objective and reliable measurements. Some of the arguments of those who fear dead uniformity sound dangerously like pleas for examinations which will be sure not to give accurate or meaningful measures — as if progress depended upon ignorance of the facts, and as if accurate knowledge of defined achievement would necessarily fasten a uniform mold upon everything pertaining to modern language work. It is an ironical paradox that the charge of dead uniformity should be made against a type of examination the main purpose of which is to enable teachers to provide for individual differences, and which is the first and only device thus far proposed that seems capable of enabling the schools to make adequate and continuous adjustments to the individual needs of students. Accurate and comparable measures do not lead to uniformity, but provide a means of escape from undesirable uniformity. The development of means for comparable measurements of body temperature has not led medical doctors to make uniform prescriptions for their patients; on the contrary, it has enabled them to vary their prescriptions more adequately to meet particular individual needs.

Reconstruction of curriculum dependent upon use of good tests. — These critics can hardly desire the same chaos in examinations as now obtains in our textbooks and courses of study. Our textbooks are a veritable Tower of Babel. Under the present organization of modern language instruction a student practically begins a new language each time that the textbook is changed, and he forgets a large part of what he has already learned. A word-count [1] of sixteen widely used French first-year books showed a total vocabulary of about 6000 root words, with only 134 words common to all sixteen books! This is certainly variety, but pedagogically it is indefensible. If we should divide these sixteen books among sixteen students, and each learned his book perfectly, the conversation with all sixteen students present would be limited to 134 words and their inflected forms! If those who accuse the objective tests of entailing dead uniformity mean that they will help to wipe out this intolerable chaos, as well as to reduce the equally incredible and vicious overlapping of our classes, then we gladly plead guilty. Accurate and comparable measurements will shortly make possible the objective demon-

[1] Wood: *A Comparative Study of the Vocabularies of Sixteen French Textbooks*, The Modern Language Journal, February, 1927.

stration of the advantages of carefully constructed and articulated courses of study, and will gradually lead to the production of texts in which the increments of vocabulary and other materials from lesson to lesson and from semester to semester will be scaled in accordance with experimental findings.

***Learning* students prerequisite to *teaching* them.** — The importance of the use of a *single* adequately constructed examination for all the language (not literature) classes, at least to and including the fourth senior high school year of modern language work, can hardly be over-emphasized. The real purpose of good examinations is to give us accurate and significant information about the achievements of our students and about the results of our teaching; while we annually spend much time and money and energy giving the traditional examinations, we have really not been getting accurate information. The variable class standards and the overlappings set forth in the first part of this report show, among other things, how much we have neglected real measurement.

Learning our students is just as important as *teaching* them; indeed, it is prerequisite to good teaching. It has been suggested that if one-fourth of the energy now devoted to *teaching* were devoted to *learning* our students, the remaining three-fourths would produce ten times the good results now achieved, and would eliminate most of the ills that attend our present half-blind teaching efforts. No one can estimate the sum total of evils which are due, directly and indirectly, to our blind shooting in the dark, in terms of misdirected efforts of teachers and students, in terms of bad habits and bad attitudes now developed in both teachers and students, in terms of the despair and martyrdom of students doomed to so-called failure, and the boredom and stultification of unidentified bright minds which are kept to the mediocre pace of the average. No one knows what prodigious results might be achieved by identifying and providing adequately for these bright minds, and by freeing modern language classes from the dismal influence of unadapted students and directing them into activities for which they are competent. We are now in the position of medical doctors who treat patients *en masse*, without troubling to make individual diagnoses or checking up on the course of the malady or on the effects of the treatment. The analogy breaks down at this point, however; because no doctor boasts of the high mortality rates in his practice, but many teachers do just this under the guise of maintaining high standards.

The trial-period in educational guidance. — The lack of comparability in the measurements afforded by the traditional examinations has kept us in the dark concerning many modern language teaching and administrative problems. What students should be permitted and encouraged to study modern foreign languages? What is the least number of years

that a student may "take" a modern language with profit? The present survey of achievement in junior high schools gives us a partial answer to this last question, and at the same time points the way to a complete answer. The partial answer is that all general policies must submit to the ordinances of individual differences, because we have shown that some students learn more in one semester than many students learn in four semesters. The way indicated for a complete answer is a combination of good tests of general intelligence used as prognostic devices and trial-periods in modern language classes in which the amount and rate of progress of each individual are ascertained at least once each month, with a view to making such trial periods as short and as effective as possible.

There is nothing new in the trial-period idea. It has been used to a certain extent in the junior high schools of New York City and elsewhere. But it could not be used effectively so long as separate examinations were given to each class, and so long as uniform examinations were given only to the fourth-semester classes. Two years is quite a fraction of a boy's total life span; yet we find some students in the fourth-semester class who are below the second-semester average and who are still on trial! The essence of the trial-period is brevity; to make trial-periods short we must have accurate and *comparable* measurements of *defined* achievements.

Reasons for opposition to new-type tests. — It seems strange that the development of sound measuring devices should have been so long delayed in the modern language field, because it seems to be as easy, if not in fact easier, to construct and apply good examinations in modern language classes, through the fourth high school or third college year, as in any other field of learning. The causes of this delay are complex, and include among other things a too great sensitivity to tradition; an exaggeration of the importance of "time-serving" as a condition for and a measure of achievement; an exaggeration of the punitive function of examinations and a too careless acceptance of the theory that examinations have a miraculous power to *create* and *raise* standards; too much emphasis by modern language teachers upon individualistic goals and standards of achievement, involving a misconception of the nature and purposes of standards, and often leading to a violation of the elemental pedagogical principle of *continuity*, of going from the known to the unknown, of building new learning on a strengthening of the old, of close articulation between successive lessons and successive courses of study; too much emphasis in the early years upon literature and what is called "cultural content," and too much faith in the reality and measurableness of an ectoplasm called "spirit of the language," with a corresponding neglect of the language itself; and finally, not to mention several others, a lack of appreciation of certain phases of the psychology of learning.

Some of those who use "spirit of the language" as a conclusive argu-

ment against new types of tests have been intemperate enough to say that they were not interested in vocabulary, and have insisted on measuring the spirit of the language in students who did not yet know its "dry bones," i.e., its words. It is under this same theory that students are plunged into foreign language masterpieces, in a hasty pursuit of cultural and spiritual riches, without troubling about the "crumbs that mere words are." It is because the language itself is culturally of very high worth, and in any case prerequisite to advanced literary and cultural courses in the foreign languages, that our tests are confined to vocabulary, reading, grammar, idioms, etc., and make no pretence of measuring the "higher and imponderable realms." If we continue our present practice of trying to make unprepared students sense the native beauties of a Goethe, a Molière, a Cervantes, or a Dante in the original tongues when they have to look up every tenth word in the dictionary, and must have every other idiom or poetic expression explained by the teacher, our students will continue to leave their modern language classes with feelings toward the great masters as ardent as the feelings of many high school students who have unwillingly toiled over Shakespeare, or who have memorized Milton's sonnets as penances for rebelling against class exercises which seemed to them supreme folly. The solution is to do frankly what many institutions are now doing — give foreign literature courses in English translation, or make sure that students can read the foreign language with a certain minimum facility before they are admitted to literature courses in the original tongues.

This cannot be done by limiting literature classes to students who have "had" at least three or four years of preparatory work; this is the old time-serving conception of achievement and preparation which was born in sin and perpetuated in iniquity. Students should be admitted to modern foreign language literature courses when they are able to read the language with the requisite minimum facility, whether they have "had" it one, two, three, or four years. Students who are not minimally prepared should not be so admitted regardless of how many years of modern language work they have "taken."

What the minimum of facility in reading is which is prerequisite to successful work in foreign literature courses must be determined by experimental research. Informing research in this direction depends upon exact and comparable measurements of defined traits, abilities, and skills. Such measurements were impossible as long as we had to depend wholly upon the subjective and impressionistic old-type examinations, made up and scored by teachers who were jealous of their individual aims and proud of their unique "standards."

There can be no doubt about the necessity of sweeping away, completely and for all time, the time-serving conception which has thus far

so largely dominated the organization and administration of modern language instruction. The continuance of this indefensible iniquity would be a crime against both teachers and students. It is the direct cause of many of the worst evils attending modern language work. It puts a premium on stupidity and laziness, and penalizes intelligence and industry. To give entirely different examinations to each of the four high school classes, for example, and to refuse all first- and second-year students admission to the third- and fourth-year examinations, is nothing but gratuitously punishing the best of the students in first- and second-year classes for not being stupid enough to require four years to surpass the average of all fourth-year students. It is precisely the highest 15 or 20% of first-year students who, in spite of our almost malicious and invidious efforts to stultify and misrepresent them, will ultimately contribute 70% of the successful students in advanced courses.

Constructive usefulness the only justification for tests and examinations. — Incredible as it may seem, some teachers take this position, candidly: "Even if some students do, through good fortune in falling in with a good teacher, or through unusual intelligence and application, learn more French in one year than average students do in four years in high school, they do not deserve fourth-year credit, because they have taken it only one year; if they have done their chore early, let them mark time; they should not receive full fourth-year credit until they have served their full sentence of four years." It seems difficult to reconcile this position with any acceptable theory of education or of simple truth. Since the primary purpose of grades is a truthful and understandable description of defined achievement to the end that the schools may wisely guide the pupils' efforts to ascend the educational ladder, and direct their own efforts into the most profitable channels, we should never be led into wilful misrepresentation of the capacity or achievement of a student by considerations which regard grades as gifts, honors, moral judgments, or signs of mythical or hidden spiritual growth. If a student can read French no better than the average first-year student, let us say that by the grade that we assign him, whether he has "taken" French three or five years or not at all; and if he knows as much French as the average fourth-year student, let us indicate that by the grade assigned to him, whether he has "taken" French only a year, or a month, or not at all. "Taking" French is not the only way in which it may be profitably learned.

Teachers everywhere are beginning to appreciate the elemental necessity for comparability of school grades. They are beginning to realize the vanity and futility of maintaining so-called standards which are proudly unique and individualistic, but meaningless, because they are expressed in unknown terms. The increasing use of objective and reliable tests is making possible research which is not only genuinely scientific,

but is on a scale large enough to secure conclusions of more than local significance. The fact that these new types of tests afford measures that are comparable will make the conclusions of all research workers and teachers at least understandable, if not acceptable, to every competent teacher. Teachers are beginning to realize that the grades they assign to their students are not merely outlets for their personal ideals, educational or otherwise, or personal predilections and aversions; but that the only excuse for or purpose of grades is to convey accurate information, couched in understandable terms, which may be used by students, parents, teachers, and school administrators in the intelligent and constructive educational and vocational guidance of our young people. In so far as grades fail of this purpose, they are vain and wasteful iniquities, like signs at cross-roads which misdirect travelers and guide them into a wilderness of confusion, doubt, and despair — blots on the landscape which ought to be swept away as public nuisances.

PART TWO

THE REGENTS EXPERIMENT WITH NEW-TYPE EXAMINATIONS IN FRENCH, SPANISH, GERMAN, AND PHYSICS

JUNE, 1925

FOREWORD

The Board of Regents of the University of the State of New York wish to tender their thanks to the Carnegie Corporation for the opportunity to conduct an extended experiment with new-type tests, which the gift of five thousand dollars has afforded them. It has given them great satisfaction to have their organization used for this purpose of investigation in the important field of examination in the schools of the State.

The results of the experiments no one could predict. The material has now come in and has been examined by Professor Ben D. Wood, who has, as shown in the following pages, made certain definite interpretations. His conclusions have been reviewed by Dr. Avery W. Skinner, Director of the Division of Examinations, and Dr. Warren W. Coxe, Chief of the Bureau of Educational Measurements, in the State Department of Education, and while they are in agreement with some of them, about others they are in considerable doubt. The responsibility, however, is with Doctor Wood; and the Board of Regents and the State Department are very appreciative of his careful and scientific study, and are much gratified that he has succeeded in securing the funds necessary for the publication of this volume. It should without doubt form the basis for still further investigations and studies in this field.

Frank Pierrepont Graves

President of the University of New York
and State Commissioner of Education

I

INTRODUCTION

The State Department of Education of New York has been increasingly interested in recent years in the new developments in the field of educational measurements. The growth of this interest has been a natural evolution, in thorough consonance with the history of the State Department's examining activities. New York State, through its Regents examinations, has long been distinguished as the only State in the Union that has seriously attempted to supply the schools and colleges with educational measurements which are comparable on a state-wide basis over a series of years.

During the last decade the volume of the Regents examinations has grown so rapidly that the burden on the State Department has become formidable and threatens to become a serious administrative problem, aside from budgetary considerations. If examinations are to be scored in time to be used for constructive educational guidance of students, time and space factors are even more important than budgetary strictures. Moreover, recent studies of the Regents examination results, conducted by the State Department, or with its active coöperation, have indicated that there is considerable room for improvement in the Regents examinations, both in the matter of *accuracy* of measurements of achievement, and in the matter of maintaining uniform and meaningful *standards*.

The present experiment was undertaken by the State Department to learn what contribution objective forms of examinations might make, under actual working conditions and on a large scale, to this double problem of increasing the validity and comparability, and decreasing the costs of the Regents examinations. In the fall of 1924, Dr. James Sullivan, Assistant Commissioner for Secondary Education, invited the author of this report to coöperate in the experiment by supplying suitable tests, by assisting the State Department in supervising the administration and scoring of the tests, and by making a detailed analysis of the results and writing a report.

In collaboration with experts in the several subject matters, the writer had been working since 1922 on a series of objective or new-type collegiate placement tests, and several of these were nearing completion in 1924. It was, therefore, easy to prepare the tests in time to be printed and administered as parts of the regular Regents examinations in June, 1925.

The subject-matter specialists whose collaboration made possible the production of the tests were, for the French test, the late lamented Professor A. A. Méras and Miss Suzanne Roth; for the Spanish test, Professor Frank Callcott; for the German test, Professor C. M. Purin; and for the Physics test, Professor H. W. Farwell.

Plan of the experiment. — In June, 1925, the three-hour periods of the Regents examinations in French, Spanish, German, and Physics were equally divided between the old- and the new-type forms of questions. The old-type subjectively scored examinations were given during the first ninety minutes, and the objective tests during the second. The principals, school department heads, and teachers throughout the state were informed in detail of the plan of the experiment about six weeks before the examinations were given. Copies of the Directions to Students and examples of each form of question in the objective tests in each subject-matter were sent to all department heads in modern languages and physics, with the request that all students who were to take the Regents examinations be as thoroughly acquainted with the objective forms of questions as possible. Teachers were urged to drill their students on the details of the new-type questions by using objective questions in their weekly or bi-monthly quizzes during the remainder of the semester. This "fore-exercise" was entirely vindicated, for in the 45,000-odd new-type tests resulting from this experiment scarcely a dozen were found in which there was any evidence that the student had misunderstood the directions or had been confused by the form of the questions.

The old-type parts of these examinations were scored in the usual way by the teachers in the schools, and the usual tentative marks assigned to the students by the schools were based upon these ninety-minute old-type parts, just as though they were regular three-hour Regents examinations. The new-type parts were not scored by the teachers in the schools, but by a special staff of clerks under expert supervision in the State Department in Albany; but the principals and teachers understood that it was planned to use new-type results in the State Department's reviewing of the teachers' ratings on the old-type parts of the examinations. Thus while the teachers knew that the *tentative marks assigned by the schools* were entirely independent of the new-type results, they knew that the State Department planned to use new-type results in determining final "Regents Grades," as distinguished from tentative or local school ratings. The fact that, on account of unforeseen circumstances, the new-type results could not be entered on the Regents records in time to be made a partial basis for determining Regents credit, does not in any manner detract from the trustworthy character of the experiment. In the first place, it was not known until two months after the examinations were administered that the new-type results could not be used according to

original plans. In the second place, the failure to use new-type results in determining Regents credits for individual students has not prevented their use in determining, objectively and reliably, the comparative achievements of different schools and of individual teachers in the same school. Even if there were a few teachers so constituted intellectually and morally as to conspire with their students to defeat the purposes of the experiment, the only effect would be to make the new-type examinations seem less valid than they really are. It is quite certain, however, that such lapses were not frequent enough to vitiate appreciably the scientific validity of this experiment.

The examinations. — The old-type parts of the examinations used in this experiment were of exactly the same nature as the usual Regents examinations in these subject-matters; the only difference is that they were shorter. They were constructed, edited, printed, administered, read, and reviewed in the regular way. The old-type parts of the French and physics examinations are reproduced below on pages 200 to 207.

The new-type tests in French, Spanish, and German are parallel in form, each consisting of three Parts:

Part I is a vocabulary test of 100 multiple-choice items. Each item consists of a foreign language word, followed by five English words which are numbered. The student selects the one of these five English words which most nearly corresponds in meaning to the foreign language word, and puts its number in parentheses at the right of the item. The score is the number of correct identifications which the student makes.

Part II is a reading-comprehension test of 75 true-false items, carefully and experimentally graded in difficulty. Each item is a statement in the foreign language of an obvious truth or of an obvious fallacy. The truth or the falsity of each statement is easily within the knowledge of any high school student intelligent enough to study a foreign language, so that it is a test of ability to read the foreign language and not a test of special knowledge. The student indicates his understanding of the statements by putting a plus sign in the parentheses at the right of true statements, and zero at the right of false statements. The score is the number of correct responses diminished by the number of wrong responses. This method of scoring is used in Part II to overcome the gross effects of guessing on the part of intellectually incautious students. Since there are only two possible answers, true or false, students unable to read a single one of the statements might toss a coin and secure 50% correct answers on the average. Thus, students who mark fifty statements by pure guess would on the average mark twenty-five correctly, and twenty-five wrongly: the scores of these students, on the average, would be 25 minus 25, or zero.

Part III is a 100-item completion test of grammar, idioms, spelling,

word-order, capitalization, accents, etc., and of active vocabulary. Each item consists of a short English sentence or phrase, followed by a foreign language translation in which one, two, or three words are missing. The student writes the word or words necessary to complete the translation on the dotted line at the right of the page. The score is the number of items perfectly completed.

Only one new-type examination in each modern language for all year-classes in that language; comparability of measures. — The reader should notice at the outset one important and fundamental feature of the new-type modern language tests used in this experiment, namely, that each of them is designed to cover the whole range of achievement in high school modern language work. The old-type Regents examinations in each language include three separate and independent examinations, one for each year-class examined; but in this experiment we have used only one ninety-minute new-type examination for all year-classes in each language. This is, beyond doubt, the most important and significant feature of the experiment, because for the first time in the history of the Regents examinations it insures *comparability* between the measures of achievement in all year-classes on a state-wide basis, and it points out a way of escaping from the old time-serving conception of educational achievement which has thus far dominated our examining methods.[1]

At first thought it seems impossible that *one* ninety-minute examination could be adequate to measure the achievement of students in second-, third-, and fourth-year modern language classes; but the results of this experiment show not only that *one* examination can measure the whole range of achievement in high school modern language work, but that one examination can measure such achievement more accurately and more usefully than three old-type examinations measure it. There are several reasons why the new-type examinations accomplish this result; these will be discussed in detail later, but two may be mentioned in passing. In the first place, the foreign language materials used in them are carefully selected on the basis of objective data, such as word-counts, inventories of grammars, etc., and these materials are further checked up by experimental evidence. Elements which are too easy or too difficult, and those which do not give positive correlations with known criteria of achievement, are eliminated. Each Part of the new-type tests is carefully scaled in difficulty on the basis of actual responses of students, so that the whole range of achievement is covered by small and regular increments of diffi-

[1] The importance of comparability in educational measures can not be overestimated; it is not too much to say that most of the maladjustments in modern language classes have been due directly to the fact that thus far modern language examinations have not afforded us comparable measures of achievement which could be used in constructive educational guidance of students. So long as separate examinations, the inter-relations of which are not known, are used for each year-class, homogeneous classes and efficient exploitation of individual aptitudes of students are almost impossible.

culty. In the second place, the questions are set up in such a form that fully ninety per cent of the irrelevant activities of the old-type examinations are eliminated, thus making possible much larger samplings of foreign language materials and of performances of the students than in the old-type. Each of the new-type language tests used in this experiment includes 275 separate and independent questions, each objectively scored. Large samplings of foreign language materials, experimentally validated and scaled in difficulty, set up in objective question-forms, — these are the features which make for *accurate* and *comparable* measurements. The new-type tests used in this experiment exist in several equivalent forms, and additional equivalent Forms can be produced as needed. Only Form A of each of our modern language series was used in this experiment. These are reproduced in full below on pages 209 to 269.

The new-type physics test consists of 144 true-false statements, all of which have been experimentally verified as to validity (see page 299 below), and which are distributed over the whole field of high school physics in such a manner as to give equitable emphasis to the various topics. The score is the number of correct responses diminished by the number of incorrect responses. This test is reproduced below on pages 270 to 282.

Scope of experiment. — This experiment involves, in effect, a state-wide objective survey of student-achievement in four departments of secondary education. In addition to the measurement results obtained in this survey, other types of information of considerable significance have been made available. Each of the new-type tests included a brief questionnaire to be answered by the student covering such matters as age, sex, high school course, number of years which had been, and number of additional years which would be, devoted to modern language (or physics) study, plans for further education, probable life work, subject-matter preferences, etc. This questionnaire was answered by more than ninety per cent of the students. Taken in conjunction with the reliable and comparable measures of actual achievement afforded by the tests, these questionnaires are capable of revealing facts of great significance for the fundamental policies and administration of these departments of instruction, in particular, and of secondary education in general. But the main purpose of the experiment was to learn what contribution the new-type examinations might make to the Regents examination system in terms of reliability, validity, and comparability of educational measurements, and in terms of budgetary economies and administrative conveniences. It is hoped that a complete[1] analysis of the questionnaire results may be presented in a later report. The present report will confine itself almost

[1] A preliminary study of the responses of 5000 students to the questionnaire on the new-type French test is to be published in an early issue of the *Modern Language Journal*.

entirely to an analysis of our examination results, and of the examinations themselves. Table 12 shows the number of examination returns for each class in each of the four subject-matters on which the following studies of reliability, validity, comparability, and costs of old- and new-type examinations are based.

TABLE 12

NUMBERS OF STUDENTS TAKING NEW-TYPE REGENTS EXAMINATIONS IN FRENCH, SPANISH, GERMAN, AND PHYSICS (EXCLUDING DUPLICATES AND A FEW WHOSE PAPERS WERE NOT USABLE FOR VARIOUS REASONS).[1]

French II	13,486	
III	6,741	
IV	489	
Total French		20,716
Spanish II	5,340	
III	2,467	
IV	226	
Total Spanish		8,033
German II	1,482	
III	667	
IV	127	
Total German		2,276
Total Modern Language		31,025
Total Physics		14,081
Grand Total		45,106

[1] Space limitations do not permit us to discuss several interesting features of Table 12; but we cannot forego the opportunity of calling attention to one indication of supreme importance to the whole theory of offering modern languages in secondary schools, namely, the relative sizes of the 2nd-, 3rd-, and 4th-year classes. Such large numbers of students in 1st- and 2nd-year classes can be justified only by the spiritual and intangible gains of students, because they certainly do not learn the *language*.

II

RELIABILITY AND VALIDITY OF THE EXAMINATIONS

Reliability of Old- and New-Type Examinations

By reliability of a test is meant the degree of its consistency, or of its self-agreement in measuring the same fact or series of facts two or more times. The most rigorous way in which to determine the degree to which a given form of examination agrees with itself is to administer two equivalent examinations of the given form to the same students, and calculate the correlation between the two sets of scores. It is not often feasible, however, to subject students to two full examinations of a given type; hence the most usual way of estimating the reliability of a test is to "split" it into random halves, and treat the scores on these random halves just as though they were derived from two independent and separately administered examinations. By calculating the correlation between the two sets of random-half scores, we learn the reliability of one-half of the examination in question. The statistical symbol for this reliability coefficient of one-half of a test is $r_{\frac{1}{2}\frac{1}{2}}$, and is read "r-one-half-one-half." The reliability of the whole examination, r_{11} (r-one-one), can be very closely estimated from $r_{\frac{1}{2}\frac{1}{2}}$ by means of the Spearman-Brown formula, $r_{11} = \frac{2\, r_{\frac{1}{2}\frac{1}{2}}}{1 + r_{\frac{1}{2}\frac{1}{2}}}$. There is some evidence that this formula overestimates the reliability of old-type subjective examinations, and underestimates [1] that of some of the new-type objective forms of tests; but the total margin of possible error in our use of this formula is probably very small in relation to the large differences between the estimated reliability coefficients of the old- and new-type examinations with which we are concerned.

Reliability of old-type examinations. — In order to make our estimate of the reliability of the old-type as favorable as possible to the old-type, only papers from the largest half-dozen schools (in terms of modern language enrolment) in the state were used for statistical study. The reading of the old-type papers in the large schools is usually done by consultation of several teachers, under the immediate supervision of depart-

[1] See, e.g., Wood: "Studies of Achievement Tests," *Journal of Educational Psychology*, January, February and April, 1926; "Measurement of Law School Work, I and II," *Columbia Law Review*, March, 1924, and March, 1925; "New-Type Examinations in the College of Physicians and Surgeons," *Journal of Personnel Research*, October and November, 1926.

ment heads of many years' experience in teaching and in reading Regents examinations; it is therefore almost certainly better than average reading. No other basis of selection was used; the schools having been selected, *all* the papers from them, both passing and failing, were used. The variability of scores in the whole group of papers from these schools turned out to be only negligibly different from that of all the schools in the state; the only noticeable difference was in the fourth-year classes. It should be noted that only the papers "claimed," i.e., passed by the schools (or very near passing in some cases), were reviewed by the State Department in Albany; the papers of students failed by the schools came to us direct from the schools. Review ratings were used for all questions in all papers which were so rated; that is, our study of the reliability of the old-type is a study of the end-product of the Regents system, not of school ratings. School ratings were used only when they were accepted, or not changed by the State Department reviewers.

For calculating reliability the old-type papers were "split" in such a way as to include exactly fifty per cent of the credits in each half, and at the same time to include each kind of question, in so far as possible, in each half. For example, the French II paper, consisting of seven questions (see page 200, below), was divided as follows:

First half:

Q. 1, translate 12 lines of French into English	25 credits
Q. 3, write tenses of three of six given verbs	15 "
Q. 6, translate into French five of seven English phrases	10 "
	50 credits

Second half:

Q. 2, translate six lines of English into French	20 credits
Q. 4, complete five of seven French phrases	5 "
Q. 5, insert verb in five French phrases	5 "
Q. 7, write a composition in French, about 75 words	20 "
	50 credits

(NOTE: Credit for oral work may be substituted for Question 7)

Each of these halves thus represents a fairly typical forty-five-minute old-type modern language examination, just as all seven of the questions represent a fairly typical ninety-minute examination. The second-, third-, and fourth-year examinations in French and in Spanish were divided in this same manner, and correlations were calculated with the results shown in Table 13. In view of the large number of cases involved, the coefficients for the second-year papers may be accepted with confidence as fairly representative of the highest expectancy of reliability in old-type

modern language examinations. These results are in agreement with all other available evidence, that the reliability of a ninety-minute old-type examination is not above 0.80. The results for the fourth-year papers should be interpreted in the light of the small number of cases involved.

TABLE 13

RELIABILITY COEFFICIENTS OF OLD-TYPE 90-MINUTE REGENTS EXAMINATIONS IN FRENCH AND SPANISH OF JUNE, 1925. THE FIRST COLUMN HEADED $r_{\frac{1}{2}\frac{1}{2}}$ SHOWS THE CORRELATIONS BETWEEN RANDOM HALVES OF THE OLD-TYPE TESTS; THE NEXT COLUMN SHOWS THE RELIABILITY COEFFICIENTS OF THE WHOLE TESTS AS ESTIMATED BY MEANS OF THE SPEARMAN-BROWN FORMULA. THE THIRD COLUMN SHOWS THE NUMBER OF CASES ON WHICH THE CORRELATIONS ARE BASED. THESE COEFFICIENTS ARE BASED ON REVIEW RATINGS OF PAPERS FROM FIVE OF THE LARGEST SCHOOLS IN THE STATE.[1]

	$r_{\frac{1}{2}\frac{1}{2}}$	S–B r_{11}	N
French II	.649	.788	1,105
III	.584	.738	867
IV	.262	.415	85
Spanish II	.565	.722	1,016
III	.538	.700	629
IV	.532	.695	95

Reliability of new-type examinations. — Table 13 should be compared with Table 14, which shows similarly derived reliability coefficients for the new-type tests. The new-type tests were split by purely random divisions, the first half of each test being made up of the odd-numbered items, and the second of the even-numbered items. Papers were chosen from schools of all sizes and classes, so that the reliability indications of Table 14 represent what we might expect of the new-type tests under average conditions. The groups of students represented in Table 14 are representative of their respective classes as to average achievement and variability of scores.

Table 14 shows the reliability coefficients of each Part of each modern language test as well as of each whole ninety-minute new-type test.

In order to compare the reliabilities of old- and new-type tests of equal time-allowances, therefore, we must confine our attention to the last three columns but one in Table 14.

The highest Spearman-Brown reliability coefficient for an old-type

[1] It was not found possible, within our time-limits, to secure a sufficient number of cases of the old-type German examinations to calculate correlations. Such data as we have lead us to believe that German examination reliabilities are not very different from those of French and Spanish examinations. We know from other studies that the reliability of a 90-minute old-type Physics examination is about 0.70, and certainly not above 0.75. (See Wood, *Measurement in Higher Education*, World Book Co., Yonkers, N. Y., 1923. Cf. esp. ch. 9.)

TABLE 14

RELIABILITY COEFFICIENTS OF NEW-TYPE REGENTS EXAMINATIONS OF JUNE, 1925, BY PARTS AND WHOLES, BASED ON RESULTS FROM GROUPS OF SCHOOLS SELECTED AT RANDOM. THE COLUMNS HEADED $r_{\frac{1}{2}\frac{1}{2}}$ SHOW CORRELATIONS BETWEEN RANDOM HALVES OF THE TESTS; COLUMNS HEADED "S–B r_{11}" SHOW WHOLE-TEST RELIABILITY COEFFICIENTS AS ESTIMATED BY THE SPEARMAN-BROWN FORMULA; COLUMNS HEADED $\sigma_{\frac{1}{2}}$ SHOW AVERAGES OF STANDARD DEVIATIONS ON THE RANDOM HALVES OF THE TESTS; THE COLUMN HEADED "N" SHOWS THE NUMBER OF CASES.

YEAR-CLASSES	PART I VOCABULARY 25 MINUTES			PART II TRUE-FALSE 20 MINUTES			PART III GRAMMAR-COMPLETION 45 MINUTES			WHOLE TEST 90 MINUTES			
	$r_{\frac{1}{2}\frac{1}{2}}$	S–B r_{11}	$\sigma_{\frac{1}{2}}$	$r_{\frac{1}{2}\frac{1}{2}}$	S–B r_{11}	$\sigma_{\frac{1}{2}}$	$r_{\frac{1}{2}\frac{1}{2}}$	S–B r_{11}	$\sigma_{\frac{1}{2}}$	$r_{\frac{1}{2}\frac{1}{2}}$	S–B r_{11}	$\sigma_{\frac{1}{2}}$	N
French II	.853	.921	7.50	.751	.858	8.88	.920	.959	10.60	.914	.955	25.24	437
III	.751	.858	5.12	.569	.726	4.90	.741	.851	5.61	.850	.919	13.09	470
IV	.752	.859	5.58	.562	.720	4.64	.742	.852	6.29	.830	.907	14.21	446
All French Classes	.860	.925	7.18	.729	.844	7.16	.893	.944	8.85	.928	.962	21.20	1353
Spanish II	.751	.858	6.67	.576	.731	5.60	.949	.975	7.96	.907	.951	17.45	480
III	.701	.825	4.12	.391	.561	4.42	.809	.895	6.89	.882	.937	12.65	355
IV	.589	.742	3.43	.508	.674	3.77	.858	.924	6.18	.786	.880	10.66	226
All Spanish Classes	.850	.919	7.02	.517	.682	5.64	.904	.950	8.99	.932	.965	19.25	1061
German II	.840	.914	6.96	.601	.751	6.44	.916	.956	9.69	.913	.954	20.22	366
III	.846	.916	5.29	.588	.741	5.46	.914	.950	9.48	.905	.950	17.05	618
IV	.738	.849	3.42	.597	.748	4.32	.770	.870	7.84	.857	.923	12.80	83
All German Classes	.855	.922	6.81	.669	.802	6.27	.898	.947	10.04	.936	.962	20.18	1067
Physics (90 minutes, 144 true-false statements)										.751	.863	13.61	575

examination is that for French II, 0.788. The highest Spearman-Brown coefficient for a new-type examination based on returns from a single year-class is 0.955 (French II); the average of new-type reliabilities based on single year-class samplings is 0.94. This is about 0.15 higher than the highest old-type reliability. The Spearman-Brown reliability of each of the new-type modern language tests, when based upon second-, third-, and fourth-year classes taken together, is above 0.96.

The reliability coefficients for Part III of the new-type tests are all higher than the highest old-type reliability. The lowest Part III coefficient is that for French III, 0.851. This means that the lowest reliability indication we have for a forty-five-minute new-type test is about 0.06 higher than the highest indication we have obtained for an old-type ninety-minute examination. This is also very nearly true of the twenty-five-minute new-type vocabulary test. If we compare old- and new-type reliabilities class by class we find that, without exception, the reliabilities of the twenty-minute true-false reading tests are higher than those of the ninety-minute old-type examinations!

Reliability coefficients of alienation. — These differences are so large that comment is unnecessary; but those who are not familiar with the mathematics of correlation coefficients are very likely to underestimate the real magnitude of the differences. The very natural assumption which non-technical readers will unconsciously make is that correlation coefficients are like any other measures such as of length or weight, and indicate merely numbers of *equal units* or of *equal amounts* of relationship. Unfortunately, this is not the case. The *numerical* difference between coefficients of 0.50 and 0.60 is 0.10, which is equal to the *numerical* difference between coefficients of 0.80 and 0.90; but the latter difference, in terms of real correlation, is about two and one-half times as great as the former.

In order to render numerical comparisons valid, we must transmute our coefficients of correlation into coefficients of alienation, in which numerical differences are always equal to real differences in degrees of relationshio or of correlation. When the ranges of talent or achievement on which the original correlations are based are equal, as in the present case, numerical comparisons of coefficients of alienation are just as valid as in the case of height and weight measurements. The coefficient of alienation (k in statistical parlance) is the ratio of an error of estimate to the error of pure chance; that is, it tells us how nearly, for example, the scores on a test approximate a *pure chance* relationship to the scores on another test, or to the facts which it is supposed to measure. Hence a numerically high coefficient of alienation means a close approximation to chance, and therefore a low correlation; a numerically low coefficient of alienation means a high correlation. Using the formula $k = \sqrt{1 - r^2}$, we have trans-

muted the Spearman-Brown reliability estimates of whole ninety-minute tests (Tables 13 and 14) into coefficients of alienation, with the results shown in Table 15.

TABLE 15

RELIABILITY COEFFICIENTS OF ALIENATION OF OLD- AND OF NEW-TYPE REGENTS EXAMINATIONS BASED ON SPEARMAN-BROWN RELIABILITY ESTIMATES OF TABLES 13 AND 14. THE THIRD ROW OF FIGURES SHOWS THE RATIOS OF OLD-TYPE TO NEW-TYPE COEFFICIENTS. THE OLD- AND NEW-TYPE TESTS HERE COMPARED ARE ALL 90-MINUTE TESTS.

	FRENCH			SPANISH			GERMAN			PHYSICS
	II	III	IV	II	III	IV	II	III	IV	
Old Type	.616	.674	.910	.692	.714	.720				
New Type	.297	.394	.421	.308	.349	.474	.296	.307	.383	.505
ratio: $\frac{O}{N}$	2.07	1.71	2.16	2.24	2.04	1.52				

The general conclusion from Table 15 is that the new-type tests are from one and one-half to two and one-quarter times as reliable as the old-type ninety-minute tests.

Standard errors of estimate. — The reader will have noticed in Table 13 (*q.v.*) that the reliability coefficients seem to decrease in magnitude almost consistently from second- to fourth-year classes. It is important to ascertain whether this decrease in the numerical magnitude of coefficients indicates a genuine decrease in reliability, or is merely a false appearance; because it involves the question of the adaptability of new-type examinations to the measurement of achievement in advanced as opposed to elementary courses. There are some eminent modern language teachers who strongly advocate the use of new-type tests in elementary modern language work, but who are unconvinced of their usefulness in advanced courses. If the reliability of the new-type tests is found to be lower in fourth-year than in second- or third-year classes, the adaptability of objective forms of examinations for advanced courses would at least be thrown in question; if, however, it is found that the reliability is as great in fourth- as in second-year classes, then doubt on this point is removed in so far as reliability is concerned, and in so far as fourth-year high school work may be considered as *advanced.*

As indicated above, correlations expressed as coefficients of alienation may be directly compared numerically when the variabilities of the groups of students involved are known to be equal, as in the case of Table 15. In order to compare correlations based on test results from student-

groups of unequal variabilities, they must first be expressed in terms of the same units, or in terms of units whose relations are known. Since in the present instance we desire to compare the reliabilities of the same new-type tests applied to three different year-classes, we may use the scales of raw score-points of the tests as the common denominators of all the correlation coefficients for each of our modern language tests. If the values of the score-points are equal throughout the whole range of scores, this method of comparison is entirely valid. There is considerable evidence, too complex for brief exposition, indicating that the score-points in each of the new-type tests are equal, or very nearly equal, in all ranges of scores.

Referring to Table 14, the reader will notice that the figures in the columns headed $\sigma_{\frac{1}{2}}$ decrease from second- to fourth-year classes even more notably than do the reliability coefficients in the columns headed $r_{\frac{1}{2}\frac{1}{2}}$ and S–B r_{11}. Thus the $r_{\frac{1}{2}\frac{1}{2}}$'s of Part II of the French test decrease from 0.751 for French II to 0.562 for French III, while the average standard deviations of the halves of Part II, the $\sigma_{\frac{1}{2}}$'s, decrease from 8.88 to 4.64! Using the formula $\sigma_{\frac{1}{2}\cdot\frac{1}{2}} = \sigma_{\frac{1}{2}} \sqrt{1-r^2_{\frac{1}{2}\frac{1}{2}}}$, we find that the standard error of estimate of one random half of Part II from the other half is 5.86 score-points for French II and only 3.84 score-points for French IV! In other words the Part II $r_{\frac{1}{2}\frac{1}{2}}$ of 0.562 for French IV indicates a degree of reliability about 40% higher than that indicated by the $r_{\frac{1}{2}\frac{1}{2}}$ of 0.751 for French II. Similar calculations have been made with all the other $r_{\frac{1}{2}\frac{1}{2}}$'s and $\sigma_{\frac{1}{2}}$'s in Table 14 with the results shown in Table 16. Like coefficients of alienation in Table 15, the standard error of estimate in Table 16 for any one of the modern language tests may be compared just as ordinary linear measurements. If we care to make the very reasonable inference that the score-points of all three modern language tests are approximately equal, then comparisons may be made between all the figures of Table 16, except those for physics. As in the case of the coefficient of alienation, the lower the numerical value of the standard error of estimate the higher is the reliability indicated.

The general conclusion from Table 16 is that the reliability of the new-type modern language tests tends to be greater for fourth- and third-year classes than for second-year classes. The only Part of the new-type examinations which shows any tendency to be less reliable as we go from the second to the fourth year is Part III. This Part is the grammar-completion form of test which is so widely used in old-type examinations. Parts I and II of the new-type examinations — the only Parts not universally sanctioned by old-type examination practices — are consistently and progressively more reliable from second- to fourth-year classes. It is rather curious that the only form of question in the new-type tests which is regularly used by old-type examination makers, even in ad-

TABLE 16

STANDARD ERRORS OF ESTIMATE OF RANDOM HALVES OF NEW-TYPE REGENTS EXAMINATIONS OF JUNE, 1925, BASED ON $r_{\frac{1}{2}\frac{1}{2}}$'S AND STANDARD DEVIATIONS IN TABLE 14, SHOWING THAT, EXCEPT IN PART III, RELIABILITY TENDS TO INCREASE AS WE GO FROM SECOND- TO FOURTH-YEAR CLASSES.

	STANDARD ERRORS OF ESTIMATE BY THE FORMULA $\sigma_{\frac{1}{2}.\frac{1}{2}} = \sigma_{\frac{1}{2}} \sqrt{1 - r^2_{\frac{1}{2}\frac{1}{2}}}$			
	Part I	Part II	Part III	Whole Test
French II	3.92	5.86	3.70	10.20
III	3.61	4.03	3.77	6.90
IV	3.68	3.84	4.21	7.92
Total French	3.67	4.90	4.00	7.90
Spanish II	4.40	4.58	2.52	7.22
III	2.91	4.07	4.05	5.96
IV	2.77	3.30	3.18	6.59
Total Spanish	3.70	4.82	3.84	6.98
German II	3.78	5.15	3.88	8.25
III	2.82	4.43	4.05	7.44
IV	2.26	3.42	5.00	6.60
Total German	3.53	4.72	4.57	7.62
Physics				8.98

vanced college courses, is the only form of test that seems less reliable for third- and fourth- than for second-year high school classes. Even for Part III, however, the differences are neither consistent nor excessively large, so that this indication should be accepted only tentatively. Moreover, these facts apply only to reliability, and reliability is by no means the ultimate criterion in judging the acceptability of given forms of tests for given purposes. The constructive indication here, that Parts I and II and the wholes of the new-type tests are equally or more reliable for fourth- and third-year classes than for second-year classes, is more important than the small reverse differences noted for Part III.

Probable errors of estimate.— Table 17 shows the probable errors of estimate of each new-type test. The figures for the three modern language tests are based on returns from second-, third- and fourth-year classes taken together. Although these probable errors are satisfactorily small, the reader should remember that they apply to halves of the new-type tests. For example, the first entry in column 1 tells us that the probable error of estimate of scores on one $12\frac{1}{2}$-minute French vocabulary test from scores on a similar $12\frac{1}{2}$-minute vocabulary test is 2.48 score-points (or words). The probable error of estimate of scores on one

ten-minute true-false French reading test from scores on a similar test is 3.30 points. The probable error of estimate of one-half (45 minutes) the whole new-type French test from the other half is 5.33 points.[1] The probable error of estimate of a score on a ninety-minute new-type test from estimated "true" scores would be about 4.45 score-points.

TABLE 17

PROBABLE ERRORS OF ESTIMATE AND OF PLACEMENT OF RANDOM HALVES OF NEW-TYPE TESTS BASED ON $r_{\frac{1}{2}\frac{1}{2}}$'S AND $\sigma_{\frac{1}{2}}$'S IN TABLE 14. THE FIRST FOUR COLUMNS SHOW PROBABLE ERRORS OF ESTIMATING SCORES ON ODD-NUMBERED ELEMENTS FROM SCORES ON EVEN-NUMBERED ELEMENTS; THE FIFTH COLUMN SHOWS PROBABLE ERRORS OF ESTIMATE OF SCORES ON ONE-HALF THE TEST FROM "TRUE" SCORES. THE FIGURES IN PARENTHESES IN COLUMN 5 SHOW THE PER CENT THAT EACH P.E.$_{\infty\,\frac{1}{2}}$ IS OF THE DIFFERENCE BETWEEN NEW-TYPE HALF-TEST AVERAGE SCORES OF SECOND- AND THIRD-YEAR STUDENTS.

	PROBABLE ERRORS OF ESTIMATE $P.E._{\frac{1}{2}\frac{1}{2}} = .6745\,\sigma_{\frac{1}{2}}\sqrt{1 - r^2_{\frac{1}{2}\frac{1}{2}}}$				PROBABLE ERRORS OF PLACEMENT $P.E._{\infty\,\frac{1}{2}} = .6745\,\sigma_{\frac{1}{2}}\sqrt{r_{\frac{1}{2}\frac{1}{2}} - r^2_{\frac{1}{2}\frac{1}{2}}}$
	Part I	Part II	Part III	Whole Test	Whole Test
French . .	2.48	3.30	2.70	5.33	3.68 (19.35%)
Spanish . .	2.50	3.25	2.59	4.71	3.29 (15.65%)
German . .	2.38	3.18	3.08	5.14	3.34 (19.0%)
Physics . .				6.06	3.97

Probable errors of placement. — The last column of Table 17 shows the probable error of estimate of "true" scores on one-half of each new-type test from obtained scores on one-half of each test. From these figures we learn that the probable error of placement of students on the basis of one-half the new-type French test is 3.68 score-points, which is 19.35% of the difference between the average scores of second- and third-year French classes. In other words, if the French students in New York high schools were reclassified on the basis of scores on a random half of the new-type French test, half of the students would have been misplaced by less than one-fifth of a year's growth in French achievement (if difference between second- and third-year average scores may be called a year's growth), and half would be misplaced by more than one-fifth of a year's growth. Less than 10% of the students would be misplaced by more than one semester, and less than 1% of the students

[1] [By taking the S–B estimate of reliability of the whole French test, and using the standard deviation of total scores of all three classes taken together, the probable error of estimate of scores on the ninety-minute new-type French test from scores on another similar test is 6.24 score points.]

would be misplaced by a whole year. This error, large as it is, is still much smaller than that of the present classification of modern language students. As we shall see (below, page 157) an average of about 30% of second- and third-year students in all three languages are misplaced by a whole year or more. The error of placement on the basis of a whole ninety-minute new-type test would, of course, be still smaller.

Validity of New-Type Tests

Thus far we have been mainly concerned with the reliability of tests, that is, their consistency in measuring whatever they measure. Reliability is an important requisite; for unless a test agrees with itself when applied to the same facts two or more times, it cannot be said to measure any definite or definable thing or function, good or bad. Other things being equal, the most desirable form of test is the one which has the highest reliability coefficient. But the most important feature of a test is *what* it measures. We have shown that the new-type tests are about twice as reliable as old-type tests of equal time-allowance, and that they are at least as reliable for fourth- as for second-year classes; but unless the new-type tests measure real modern language achievement as well as, or better than, the old-type examinations do, no degree of reliability or consistency in measuring something else would make them suitable additions to the present Regents modern language examinations. We have two or three ways of arriving at some conclusion as to *what* the new-type tests actually measure, — (*a*) their correlations with the old-type parts of the Regents examinations in the same subject-matters, (*b*) the intercorrelations of the three Parts of the new-type language tests, and (*c*) analysis of the internal constitution of old- and of new-type tests. We shall here take up (*a*) and (*b*), reserving (*c*) for a later section of this report (pages 198–307). In considering the correlations between old- and new-type tests set forth in Table 18, the reader must recall the very low reliability coefficients found for the old-type tests and displayed in Table 13. Whatever the old-type examinations measure, they measure it so unreliably and inconsistently that no test, however valid and reliable, could correlate with them very highly, except by chance.

Correlations of new- with old-type examinations. — In view of the large numbers of cases involved, the correlations of Table 18 may be accepted with considerable confidence. Of the twenty correlations in the table, four are based on more than 13,000 cases, eight on more than 5,000 cases, and twelve on more than 1,000 cases each. The correlations of new-type with school ratings, except those for French IV and physics, range from 0.64 to 0.71. Considering the varying standards of severity of marking in different schools in the state, and the lowering effect which this mixing

of standards necessarily has on relationships between the new- and old-type tests, these correlations are satisfactorily high. That the correlations would have been higher if all the papers had been marked by the same readers is indicated by the fact that all but two of the correlations of new-type with Regents review ratings (column 2) are higher than the corresponding correlations with school ratings (column 1). The review ratings

TABLE 18

VALIDITY COEFFICIENTS AND STANDARD ERRORS OF ESTIMATE OF NEW-TYPE REGENTS EXAMINATIONS OF JUNE, 1925. THE CORRELATION OF NEW-TYPE WITH REVIEW RATINGS IS HIGHER THAN WITH SCHOOL RATINGS IN ALL CASES EXCEPT SPANISH II AND III. THE STANDARD ERRORS SHOW THAT THE VALIDITIES OF THE NEW-TYPE TESTS IN SPANISH AND IN GERMAN BECOME GREATER AS WE GO FROM SECOND- TO FOURTH-YEAR CLASSES, WHILE THE FRENCH TEST VALIDITIES ARE FAIRLY CONSTANT. THERE IS NO INDICATION ANYWHERE IN OUR DATA THAT THESE OBJECTIVE TESTS BECOME LESS ADEQUATE AS WE GO UP THE SCALE OF ACHIEVEMENT IN MODERN LANGUAGES. ON THE CONTRARY, THEY BECOME BOTH MORE RELIABLE AND MORE VALID.

	Correlations of New-Type with		Standard Errors of Estimate of New-Type Scores from Old-Type School and Review Ratings		N
	School ratings	Review ratings	$\sigma_{N.S}$	$\sigma_{N.R}$	
French II	.686	.710	23.0	22.2	13,486
III	.671	.724	20.4	18.9	6,741
IV	.577	.579	22.3	22.3	489
Spanish II	.699	.551	22.1	25.7	5,340
III	.670	.654	18.7	19.1	2,467
IV	.696	.713	14.7	14.5	226
German II	.669	.719	29.1	27.2	1,482
III	.711	.756	23.6	22.0	667
IV	.640	.723	19.5	17.5	127
Physics	.556	.582	19.0	18.6	14,081

of Spanish II and III papers give correlations with new-type scores of 0.551 and 0.654, respectively, while the school ratings give corresponding correlations of 0.699 and 0.670. We have no satisfactory theory to offer in explanation of these differences. The difference for Spanish III is small, and may be ignored as due to chance factors; but the difference for Spanish II is both large and genuine, for the correlations here are each based on over 5,000 cases. Six of the correlations in column 2 are greater than 0.70, which is as strong a vindication of the validity of the new-type tests as we could well expect from the old-type examinations here used as criteria.

New-type tests more valid for 4th- and 3rd- than for 2nd-year classes. — In preceding pages we have shown that the new-type tests are at least as reliable for fourth-year classes as for second-year classes. The question of the validity of objective tests for advanced modern language work was left to later discussion. The coefficients in columns 1 and 2 are lower for fourth- than for second-year classes, and thus superficially bear out the position taken by a few modern language teachers that while objective tests are very useful for measuring achievement in elementary courses they are not so well adapted to the measurement of the higher aspects of modern language work in advanced courses. But the raw correlations here are as deceptive as in Table 14 above. The standard errors of estimate in columns 3 and 4 in Table 18 show that the new-type test is equally valid in all three French classes, and is consistently and progressively more valid as we go from second- to fourth-year classes in Spanish and German. This is an important finding, and ought to be convincing to the most devoted believers in the capacity of the old-type to reveal the higher products of modern language teaching, because the superior validity of the new-type for fourth-year classes is here set forth in terms of very much closer correlation of new-type with fourth-year old-type examinations than with second- and third-year old-type examinations.

Correlation of Parts of new-type tests with old-type examinations. — Table 19 shows that, in terms of agreement with old-type examination results, each of the three Parts of the new-type tests is satisfactorily valid. The forty-five-minute grammar-completion test gives higher coefficients than the twenty-five-minute vocabulary and twenty-minute reading tests, as was to be expected. The most significant indication of Table 19 is that Parts I and II become more valid as we go from second- to fourth-year classes, while Part III either remains about the same or actually becomes less valid in the higher classes than in the lower. The standard errors of estimate of Part I and II scores from old-type scores consistently become smaller from second- to fourth-year classes, with the single exception of the French vocabulary test, while the standard errors of estimate of the grammar-completion Part either remain fairly constant or increase. These facts, with those presented in preceding pages, constitute a conclusive vindication, *in terms of strictly old-type criteria*, of all three new-type Parts, but particularly of the two most suspected Parts, the vocabulary and the reading tests. If the new-type is valid for second-year classes, which few would doubt even on *a priori* grounds, it is at the very least just as valid for fourth-year high school modern language work. If the reputed higher and less tangible values of fourth-year work really exist, and are measured by the old-type tests here used as criteria, the conclusion is inescapable that the objective tests of vocabulary and reading catch these elusive qualities at least as effectively as they measure the

less dignified achievements in second-year work. The most likely explanation is that fourth-year achievement is not different in fundamental nature or worth from second-year achievement, and that most or all of the greater mystery and intangibleness of the former is a pleasing but fictitious importation. Genuine progress in advanced modern language work depends quite as much on effective reading ability and knowledge of vocabulary as in elementary courses, with the difference that more of both is desirable. Those who decry the "mere mechanics" of the

TABLE 19

VALIDITY OF PARTS OF REGENTS NEW-TYPE MODERN LANGUAGE TESTS. CORRELATIONS BETWEEN REGENTS OLD-TYPE 90-MINUTE EXAMINATIONS AND EACH PART OF NEW-TYPE TESTS, WITH STANDARD ERRORS OF ESTIMATE OF SCORES ON EACH NEW-TYPE PART FROM SCORES ON OLD-TYPE EXAMINATIONS.

SUBJECTS AND CLASSES	CORRELATIONS			STANDARD ERRORS OF ESTIMATE			N
	Part I Vocabulary 25 min.	Part II Reading 20 min.	Part III Grammar 45 min.	$\sigma_{I.R}$	$\sigma_{II.R}$	$\sigma_{III.R}$	
French II	.448	.552	.693	8.3	8.2	7.9	543
III	.500	.313	.602	7.2	8.5	6.3	475
IV	.540	.464	.566	8.9	6.8	10.0	444
Spanish II	.684	.577	.816	8.4	8.5	8.6	492
III	.568	.394	.724	6.5	6.9	8.8	567
IV	.484	.480	.740	5.6	5.9	8.2	225
German II	.581	.556	.714	10.8	9.5	11.5	578
III	.644	.613	.541	8.2	8.1	15.0	513
IV	.533	.468	.673	5.2	6.6	11.2	83
Averages	.554	.491	.674				

language in advanced examinations are very likely guilty of sacrificing genuine progress in modern language work to a disembodied goal which, probably fortunately, has no real existence. In these new-type tests half the examination period has been allotted to vocabulary and reading; the indications are that more, not less, time should be given in all our examinations, particularly for advanced courses, to the humble but indispensable matter of vocabulary and reading.

Intercorrelations of new-type Parts. — It is a matter of fundamental importance to the theory and practice of modern language examining to know the extent of the interdependence of the three supposedly distinct aspects of modern language achievement represented by the vocabulary, reading, and grammar-completion Parts of the new-type examinations. Some teachers have undoubtedly underestimated the importance of range of vocabulary as a condition for, and measure of, modern language achieve-

TABLE 20

INTERCORRELATIONS OF PARTS OF NEW-TYPE MODERN LANGUAGE TESTS WITH SIGMAS AND STANDARD ERRORS OF ESTIMATE, SHOWING THAT THE RELATIONSHIPS BETWEEN THE PARTS TEND TO REMAIN FAIRLY CONSTANT FROM THE SECOND TO THE FOURTH YEARS, INCLUSIVE. PART I IS THE 25-MINUTE VOCABULARY TEST; PART II IS THE 20-MINUTE TRUE-FALSE READING-COMPREHENSION TEST; AND III IS THE GRAMMAR-COMPLETION TEST OF 45 MINUTES. ASSUMING THAT THE UNITS OF THESE THREE TESTS ARE APPROXIMATELY EQUAL THROUGHOUT THE WHOLE RANGE OF SCORES, THIS TABLE INDICATES AMONG OTHER THINGS THAT EACH OF THE THREE PARTS IS AS VALID FOR FOURTH-YEAR HIGH SCHOOL CLASSES AS FOR SECOND-YEAR CLASSES.

	Correlations between Parts			Standard Deviations			Standard Errors of Estimate						N
	I–II	I–III	II–III	I	II	III	$\sigma_{1.2}$	$\sigma_{1.3}$	$\sigma_{2.1}$	$\sigma_{2.3}$	$\sigma_{3.1}$	$\sigma_{3.2}$	
French II	.658	.714	.699	11.05	11.85	13.50	8.33	7.74	8.93	8.48	9.45	9.66	13486
III	.624	.652	.609	10.20	9.65	11.55	7.97	7.59	7.54	7.65	8.48	9.16	6741
IV	.550	.698	.531	10.16	8.18	11.91	8.58	7.18	6.84	6.93	8.53	10.80	489
Total French	.714	.737	.705	12.66	12.39	14.43	8.90	8.53	8.66	8.79	9.76	10.23	20716
Spanish II	.629	.702	.578	11.43	9.69	14.16	8.88	7.12	7.48	7.91	10.10	11.65	5340
III	.522	.605	.486	8.79	7.83	12.78	7.50	7.00	6.68	6.85	10.16	11.16	2467
IV	.378	.573	.417	6.57	6.81	12.15	6.09	8.20	6.31	6.14	9.96	10.95	226
Total Spanish	.654	.744	.613	12.23	9.55	15.09	9.25	8.17	7.23	7.54	10.01	11.90	8033
German II	.738	.685	.583	14.10	8.21	17.70	9.52	10.28	5.54	6.68	12.90	14.40	1482
III	.741	.730	.632	9.77	9.45	17.74	6.51	6.68	6.35	7.33	11.89	13.30	667
IV	.443	.487	.468	6.64	7.03	15.96	5.96	5.80	6.31	6.21	13.94	14.00	127
Total German	.776	.729	.653	14.24	12.27	18.80	8.98	9.75	7.72	9.31	12.32	14.25	2276

ment; indeed, deriding the value of "mere vocabulary" is a not infrequent indoor sport at some teachers' meetings. The magnitude of the intercorrelations set forth in Table 20 shows how ill-founded is the popular disrespect for word-knowledge.

The indications of Table 20 may be accepted with confidence, since most of the correlations are based on several thousand cases each. The general conclusion is that the intercorrelations of the vocabulary, reading, and grammar tests are high enough to establish the validity of each of them for general measurement purposes, and yet low enough to show that each has a distinctive contribution to make to the measurement of total achievement. It is of special interest to notice from the standard errors of estimate that the interrelations are, on the whole, closer in fourth- than in second- and third-year classes. The high correlations of the vocabulary and reading Parts with an old and tried form of examination such as the grammar-completion test are strong evidence of their validity; that their correlations are higher in advanced than in elementary high school classes is, with the facts presented in preceding pages, practically conclusive evidence of the high validity and indispensable values of the objective vocabulary and reading tests.

The reader will have noticed in preceding tables, as well as in Table 20, that the errors of estimate of the true-false reading tests have decreased more notably and more consistently from second- to fourth-year classes than the errors of estimate of any other form of test. This seems to be due in part to the greater adaptability of the true-false form of question to brighter and more advanced students than to younger and duller students. In other words, the reliability and validity of true-false tests seems to depend more on the intelligent, though perhaps unconscious, coöperation of the students than in the case of other forms of tests. It is quite likely that the multiple choice form of reading test, such as was used in the junior high school examinations,[1] is better for Regents purposes than the true-false form. The former is much easier to construct than the latter, and suffers from fewer restrictions as to content, vocabulary, and grammar. All things considered, and pending further evidence, the writer believes that multiple choice reading tests should be used to the exclusion of true-false tests except for experimental purposes. The slight tendency of the probable errors of estimate of the grammar-completion test to increase from second- to fourth-year is probably due, in part, to the variableness of the content of the fourth-year course of modern language study. The tendency of some teachers to take their students into the vast fields of foreign language literature before they have mastered the fundamental vocabulary, grammar, and idiom of the language itself is too well known to need comment; but many wise teachers do not

[1] See Part I of this volume, particularly pp. 55 to 61 and 73 to 79.

assume that fourth-year students can do without further extensive instruction in grammar. Thus we have some fourth-year students so extensively engaged with literary selections that they perforce neglect effective exercise in grammar, while other fourth-year students are assiduously strengthening and extending their active knowledge of grammar. The result is to make grammar test scores less valid as measures of *total* achievement in fourth- than in second- and third-year classes.[1]

[1] Professor Henmon suggests that "The tendency of the probable errors of estimate in grammar to increase is probably due to the differences in emphasis on grammar in different classes, at different levels of progress and with differing methods. It is very noticeable in scoring different classes that some are above the average in grammar but below in both vocabulary and reading. On the other hand, individual classes can be found which are above in vocabulary and reading and below in grammar. Pooling the results from a large number of schools obscures these differences but produces a great variability, especially evident in grammar."

III

SCHOOL DIFFERENCES: VALIDITY OF OLD-TYPE EXAMINATION RESULTS IN INDIVIDUAL SCHOOLS AND CLASSES OF SCHOOLS

New-type results used as criteria for measuring validity of old-type examination results.— Although we have not yet presented all our data tending to establish the validity of the new-type tests, we believe that sufficient evidence has been presented to warrant the conclusion that they give more valid and more comparable, as well as more reliable, measures of total modern language achievement than the old-type tests used in this experiment. For convenience of exposition in connection with Table 21, we shall assume this conclusion to be correct, in the firm belief that when all our evidence has been presented the reader will find the conclusion more than justified. The reader is therefore asked henceforth to consider the new-type results as criteria of language achievement, and to consider ratings on the old-type examinations from different schools and classes of schools and from the State Department reviewers as being on trial. Thus we shall tentatively conclude that those schools whose ratings agree most closely with new-type scores are more accurate and valid in their reading of the old-type examinations than schools whose ratings agree less closely with new-type scores.

In this report we shall call attention to only a few of the more important indications of Table 21, leaving it to the special interests of readers to examine other parts of the Table in detail.

Validity of ratings from individual schools extremely variable.— Looking first at column (2) in Table 21, it is apparent that the correlations between new-type scores and school ratings on old-type examinations vary for different schools and classes of schools. The correlations in the upper half of column (2) (French II classes) vary from 0.599 to 0.875; and those in the lower half of column (2) (French III classes) vary from 0.517 to 0.810. These variations in degree of agreement with new-type scores indicate that the accuracy of rating the old-type examinations is notably greater in some schools than in others.

Validity of school ratings increases consistently with size of schools.— Still looking at column (2), it appears that the accuracy of reading old-type examinations is greater in large schools than in small schools. The

TABLE 21

INTERCORRELATIONS BETWEEN NEW-TYPE SCORES, SCHOOL RATINGS AND REVIEW RATINGS ON OLD-TYPE EXAMINATIONS IN FRENCH FOR INDIVIDUAL SCHOOLS AND CLASSES OF SCHOOLS, SHOWING VARIATIONS IN ACCURACY OF SCORING IN DIFFERENT SCHOOLS, AND SHOWING THAT SOME SCHOOLS THAT MOST NEEDED REVIEWING ESCAPED IT. THE VARIATIONS IN CORRELATIONS ARE PARTLY ACCOUNTED FOR BY DIFFERENCES IN VARIABILITIES OF INDIVIDUAL SCHOOL GROUPS. NT STANDS FOR NEW-TYPE, S FOR SCHOOL RATINGS, AND R FOR REVIEW RATINGS.

French II

School No.	Correlations between			Sigmas			Means			N
	NT and S	NT and R	S and R	NT	S	R	NT	S	R	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
337	.754	.804	.986	25.07	9.16	9.2	152.3	75.92	75.9	102
369 (Acad.)	.652	.555	.779	38.7	3.57	2.99	78.6	65.21	64.7	58
591	.673	.665	.946	26.1	6.90	6.98	145.0	74.4	74.7	59
791	.819	.822	.973	23.5	8.26	8.0	175.0	78.37	78.3	70
806	.744	.739	.988	25.6	8.52	8.38	169.6	77.9	78.2	178
819	.723	.722	.937	22.95	8.24	8.12	146.4	72.18	71.9	144
862	.749	.783	.990	22.7	7.82	7.86	175.0	75.93	75.9	335
884	.733	.739	.997	25.02	9.22	8.90	155.0	76.12	76.2	143
919	.875	.828	.965	29.38	9.00	9.16	152.1	75.85	76.2	129
696	.725	.724	.990	22.96	10.00	10.75	166.7	79.55	79.3	132
942	.771	.769	.991	25.58	8.06	8.3	162.6	75.22	75.3	62
971	.751	.765	.988	23.37	8.74	8.94	181.7	74.98	74.9	241
Average 50+	.747	.744	.961	(25.00)						
30–50	.714	.708	.969	31.3	8.11	8.32	147.76	72.99	72.94	1426
15–30	.713	.719	.946	31.7	7.75	8.8	145.18	71.89	71.86	2230
1–15	.700	.675	.921	30.9	5.66	8.32	146.09	72.21	71.29	2786
Sr. Schools	.649	.617	.790	31.6	6.9	5.96	137.12	70.89	69.36	358
Academies	.599	.687	.863	33.9	8.12	8.8	144.68	72.47	71.23	1228
Total Fr. II	.686	.710		31.55	7.96	8.38	150.4	73.54	72.74	13,486

French III

337	.779	.781	.973	18.75	7.51	7.74	196.8	80.97	80.79	61
696	.751	.719	.989	22.55	9.45	9.72	192.9	77.44	78.19	68
701	.749	.721	.937	21.60	8.04	7.8	196.8	77.0	76.9	48
791	.810	.819	.969	17.53	9.18	8.88	192.6	77.82	78.4	45
806	.739	.713	.939	27.3	7.98	9.02	187.4	78.06	77.3	66
819	.703	.707	.981	22.75	7.88	7.89	188.7	72.23	72.7	43
862	.517	.511	.996	17.10	7.325	7.22	196.5	79.17	79.3	235
881	.685	.668	.914	22.94	8.78	8.91	178.0	79.22	78.8	46
883	.723	.759	.872	22.10	8.88	7.02	177.6	75.44	74.09	43
884	.591	.617	.983	21.49	9.16	9.04	192.5	80.57	80.7	56
919	.753	.733	.989	22.50	8.3	8.4	183.0	74.53	74.4	57
971	.688	.656	.977	18.5	7.16	8.01	201.1	76.68	76.1	118
984	.680	.689	.994	20.0	8.09	8.06	194.6	78.72	78.8	263
Average 40+	.705	.699	.963	(21.6)						

TABLE 21 (*Continued*)

30–40	.743	.760	.978	25.77	8.39	8.39	184.89	74.88	74.77	460
15–30	.702	.698	.943	25.90	8.58	8.57	182.84	75.44	74.46	749
1–15	.652	.653	.900	26.4	7.84	7.82	182.1	74.25	72.23	1589
Sr. Schools	.582	.537	.789	31.32	6.93	6.46	163.79	73.39	69.39	58
Academies	.647	.657	.833	35.64	7.95	7.26	166.35	73.14	70.48	728
Total Fr. III	.671	.724		27.42	8.24	8.58	183.44	75.67	74.63	6741

French IV

Total Fr. IV	.577	.579		27.36	8.57	9.01	205.84	77.89	76.31	489

schools studied individually in Table 21 were selected so as to be representative of large schools all over the state, from among those having fifty or more students in French II, and forty or more in French III. The identifications by numbers of the individual schools and classes of schools are given in column (1). The correlations for the largest schools have been averaged, so that we may compare the schools that have fifty or more students in French II or III, as a group, with groups of schools having 30–50, 15–30 and 1–15 students, and with senior schools [1] as a group, and with the academies (private schools) as a group. The average correlation between new-type and school ratings on the French II old-type examination for the largest schools is 0.747, and for the academies only 0.599. The decrease in correlation as we go from the large schools through the smaller schools and senior schools to the academies is consistent.

Validity of ratings in such high schools as Morris, Boys, Albany, Utica, South Park, etc., nearly twice as great as in academies. — The differences between the six size-groups of schools studied in Table 21, as to accuracy of old-type examination ratings, are much greater than is indicated by the correlation coefficients in column (2). The apparent reversals in the order of correlations at the bottom of column (2) also disappear entirely when the correlations are interpreted in relation to the standard deviations of the groups on which they are based. Table 21*a* shows these relationships in terms of comparable units, that is, in terms of standard errors of estimate of new-type scores from old-type ratings.

From Table 21*a* it appears that, in terms of agreement with new-type scores, the ratings of the old-type examinations in such schools as Morris, Boys, Albany, Utica, and South Park High Schools, etc., are roughly nearly twice as accurate and significant as they are in the academies and

[1] A *senior school* is defined by the Handbook of the University of the State of New York as an institution that is "registered as affording suitable facilities for maintaining an approved course of three years of academic work."

senior schools in New York State. This indication, based upon objective and unbiased data, is of great significance with regard not only to the examinations of the State Department, but to its fundamental educational administrative policies. Taken in conjunction with the uniformly lower average achievements of the smaller schools and academies, and the greater heterogeneity of their classes [Table 21, columns (5) and (8)], Table 21*a* is basic evidence of the need for a centralized examining agency which can furnish all the schools of the state with accurate measurements

TABLE 21*a*

STANDARD ERRORS OF ESTIMATE OF NEW-TYPE SCORES FROM OLD-TYPE EXAMINATION RATINGS FROM SIX SIZE-GROUPS OF SCHOOLS, CALCULATED FROM THE CORRELATIONS AND SIGMAS OF TABLE 21, BY THE FORMULA $\sigma_{N.0} = \sigma_N \sqrt{1 - r^2_{N0}}$.

SCHOOL GROUPS	FRENCH II	FRENCH III
50 + . .	17.0	15.3
30–50 . . .	22.0	16.7
15–30 . . .	22.2	18.4
1–15 . . .	22.1	20.0
Senior Schools	24.0	25.4
Academies .	27.0	27.1

of defined achievements expressed in comparable units. Abandonment or weakening of the Regents examinations would be a retrogression to chaos so far as accurate and meaningful measurements of school products are concerned.

That the large schools also need a centralized agency is indicated by the differences between the correlations for the individual schools in column (2). For example, the French III classes of schools 791 and 862 are almost equal in average achievement and in variability [columns (5) and (8)]; but the correlation of old-type school ratings with new-type scores for the former is 0.81, and for the latter only 0.517 [column (1)]. The accuracy of rating in 791 is about 50 per cent greater than that in 862, in terms of standard error of estimate. This may be due partly to the fact that in 791 there were only 45 students to rate, while there were 235 in 862. But comparing 791 with 806, which had only 66 students in French III, we find a still greater difference! The standard error of estimate in 806 is almost twice as great as in 791!

Validity of Regents review ratings varies as much from school to school as validity of school ratings varies and is only slightly greater on the whole than validity of school ratings. — The figures in columns (3) and (4) in Table 21 show that while the Regents reviewing does slightly increase the accuracy of the school ratings on the old-type papers, when the total

populations of each of the classes are treated as groups, yet the reviewing does not nearly raise the accuracy of the marking in the vast majority of the schools to the high level of accuracy achieved by such schools as those named above. Indeed, the correlations in column (3) between new-type and review ratings vary as much as those in column (2). One would not expect the reviewing to improve on the correlations of 0.75 and above, which the best schools give, because these coefficients probably represent the maximum validity of which a ninety-minute old-type examination is capable; but it is certainly within reason to expect that thorough reviewing would raise validity coefficients which are as low as 0.517 to 0.70, that is, to a figure approximating the average of the validity coefficients from the dozen schools in the state which are known to be among the most accurate. These large schools show that it is possible to reduce the standard error of estimate to an average of about 17 new-type score-points for French II, and 15 for French III; but all schools having less than 50 students in French II give standard errors of from 22 to 27 new-type score-points, and all schools having less than 40 French III students give standard errors of from 17 to 27 score-points. *These standard errors are in no case significantly reduced by the Department reviewing:* if for a given school, the school ratings give a small standard error (i.e., a high correlation) the review ratings will do the same, within a point or two; if the school ratings show a large standard error (i.e., a low correlation), the review ratings also will show a low correlation with new-type scores. Thus, the reviewing turns out to be, in effect, a fairly general acceptance of the school ratings. In some cases the review marks give correlations with new-type scores actually lower than those which the school ratings give: comparison of columns (2) and (3) shows that this happens in about a third of the cases in Table 21. In a later section of this report we shall show that the reviewing fails also to equalize the varying standards of the schools.

These facts signify a serious and genuine breakdown of the Regents examinations. They show that the state examination system is not, in fact, giving the schools uniformly and highly accurate educational measurements nor maintaining uniform standards of achievement for all the schools, but is to a large extent simply accepting the ratings of the schools, with most of the local variations in standards and accuracy unaltered, and is "giving" them the name of *Regents Grades.* What is the reason for this failure to achieve the very ends for which the Regents examinations were mainly established, and the accomplishment of which are so essential to sound educational administration and to definable progress?

Reasons for failure of Regents reviewing to correct inaccuracies of school ratings on old-type Regents examinations. — It would manifestly be absurd to charge this to the readers or to the State Administration. The

readers are all thoroughly trained as readers, have had a minimum of several years of teaching experience, and work under the immediate direction of the State Supervisor. They do their work devotedly, and take just pride in it. The real causes of the failure are not hard to find.

Not all papers are reviewed. — In the first place, not all of the papers are reviewed. The number of readers provided is so small in relation to the number of papers to be read that the State Department is compelled to rely on the "sample method" of reviewing. All the papers from a given school must be judged on the basis of a review of ten or fifteen per cent of them.

Some reviewing is misplaced. — The reviewing that does take place is frequently misplaced; that is, some papers from a given school that need reviewing may escape it, and some that do not need reviewing get it; and some schools that need extensive reviewing escape with a "fortunate" sampling, and others that do not need it have the time of the reviewers taken up needlessly because of an "unfortunate" sampling. This comes about because the sample itself is the only guide which the readers have in selecting papers and schools for review. The full quota of reviewing time may be expended on papers from a given school before the reader is enabled to make a reliable judgment as to how complete a review that particular school ought to have. The size of the sampling of papers reviewed from a given school depends on the quality of the first papers that happen to be chosen to make up the sample. As an illustration of misplaced reviewing efforts, the case of schools 791 and 862 may be mentioned again. As indicated above (page 132), the French III school ratings in the former give a correlation of 0.810 with new-type and in the latter of 0.517. If the new-type has any reasonable validity, 862 clearly needed reviewing more than 791; but the correlations between review and school ratings for these two schools are 0.969 and 0.996, respectively, showing that the school most in need of review had its ratings accepted practically without change.

Review ratings are not independent. — A third disadvantage under which the Department reviewing suffers is that the Regents system does not provide for genuinely independent ratings by the Department readers. Under the present system the Department reader can hardly avoid seeing the marks of the school readers, and it is well known that truly independent ratings are not possible under such conditions. The Department reader, however careful and conscientious, cannot entirely overcome the "drag" effect which his knowledge of the school readers' marks gives to the school ratings; the tendency, therefore, is to make changes only when the Department reader is *fully* convinced of the errors in the school ratings, which means that many small errors escape correction in papers that are reviewed. The dependent character of the Department review-

ing is a more serious fault than it appears to be, especially in conjunction with the other weaknesses here listed.

Unclaimed papers very rarely reviewed. — Except in rare cases, the Department reviewers never see the papers not *claimed* by the schools. In the ordinary routine of the Regents examinations, only papers claimed by the schools are sent to Albany for review. If the reviewers become convinced that a certain school has marked too severely, a special case is made of that school, and it is requested to send up its failed or unclaimed papers. This weakness in the Regents system might have been mentioned as an aspect of the first defect described above under the head of incompleteness; because in effect it means that school ratings on about one-fifth of all the papers are automatically accepted by the State Department, and these ratings become official Regents Grades forever after. Thus about two out of three students that are failed on the Regents examinations are failed on local standards by the schools, without benefit of Department reviews! But this aspect of the incompleteness of the Department reviewing is mentioned separately because it indicates at least a slight survival in the Regents system of the ancient punitive and unconstructive theory of examinations. The fact that only *claimed* papers are regularly sent up for review is, on its face, a tacit reversal of the legal doctrine that innocence must be assumed until guilt is proven beyond reasonable doubt. It says, in effect, that the State Department assumes that all the schools will "claim" *all* the papers which are entitled to a Regents passing mark, *or more*, and that no school will fail to claim all papers that it ought to claim. Aside from the reflection on the integrity of the school reader, this is a singularly unconstructive practice, both in theory and effect. The State Department ought to be more zealous in protecting the educational interests of good students than it is in keeping some schools from making a false showing of high achievement. The practice of reviewing only claimed papers, and that of confining the reviewing of these mainly to those just above the passing line established by the schools, *assumes* too much about local school standards which the Regents examinations were established to *verify*.

Of course, this is in no way a criticism of State Department officials. With the provisions available, the Department could hardly do more or better reviewing than it is doing. The present experiment, and others like it which the Department has carried out, show that the state educational officers are keenly aware of the weaknesses of the Regents system, and are seeking constructive and sound remedies as rapidly as possible.

Fundamental weakness inherent in old-type subjective form of examinations. — The four weaknesses thus far described have had to do mainly with the methods and conditions of the Department reviewing which keep the review ratings from being notable improvements on the school ratings

as to accuracy; but the fundamental weakness of the Regents system, from which all the defects described arise, is inherent in the nature of the old-type examination. Even the results from the most accurate schools fail to approach what might be considered as satisfactory reliability and validity coefficients, so that reviews would in no case give entirely satisfactory results no matter how extensive they might be. The old-type form of examination is subjective in both structure and scoring, it can include only very inadequate samplings of materials and performances in examination periods of normal length, it is incapable of giving comparable measures, particularly between different year-classes, and finally, as we shall show in a later section, it is extravagantly costly in comparison with other more reliable forms of examinations. The objective or new-type forms of tests used in this experiment will go far toward eliminating all of these weaknesses.

This does not mean that old-type forms of questions should be abandoned. On the contrary, our data show that the old-type examinations have considerable measurement value, when the papers are carefully and *completely* reviewed by a central agency. But in no case does it appear that the old-type is more than a partial measure of total modern language achievement. The old-type examination *at its best* gives no better indications than we have secured from the least reliable and least valid of the three Parts of the new-type examinations. There is no more reason for limiting the Regents modern language examinations to old-type questions than for limiting them to true-false reading questions or to multiple choice vocabulary questions. The two types supplement each other; the answer, therefore, is to use a judicious combination of both types. We shall thereby minimize the weaknesses and smooth out the defects of both while surrendering the real values of neither of the two types of examinations.

Classes in large schools are more homogeneous and achieve higher average scores than in small schools. — A fifth indication in Table 21 is that the classes are more homogeneous in the larger schools than in the smaller. The figures in column (5) show that the standard deviations become progressively larger as we go from schools having fifty or more students in French II and forty or more in French III, through the smaller schools to senior schools and academies. The average standard deviation for the French II classes in large schools is about twenty-five score-points and in academies almost thirty-four points. The corresponding standard deviations for French III classes are 21.6 and 35.6 score-points. Only two of the twelve schools represented in Table 21 having more than fifty students in French II have sigmas above twenty-six points; one of these (No. 369) is a private school, with standard deviation of 38.7 points, and the other is School 919 with standard deviation of 29.38 points.

Inspection of column 8 will show also that the average scores become progressively smaller as we go from large to small schools. Only three or four of the large schools are below the state-wide averages for French II and III; and of these few only one, No. 369 (a private school) is more than five points below average. The average new-type score of the French II class in School No. 369 is seventy-two points below the state-wide average.

Table 21 thus shows three important differences, which seem to be highly interdependent, between the large schools on the one hand, and the small schools and private schools on the other: (1) the large schools furnish more accurate ratings on the old-type Regents examinations than the smaller schools and academies; (2) the classification of students is better in the large schools, the classes in some small and private schools being so heterogeneous as to make efficient teaching very difficult, if not impossible; and (3) the average achievement in large schools is uniformly at or above the state-wide average, while that of the smaller and private schools is consistently and often considerably below the state-wide average. These differences hold for both second- and third-year classes, being slightly more pronounced for the latter.

Proportions of students of French in various classes of schools. — The sixth and last fact that will be specifically mentioned in connection with Table 21 is derived from column (11), which shows the numbers of students from each of five classes of schools that took the Regents examinations in June, 1925. It appears that about 9% of all the French II students in the state come from academies, 2.6% from senior schools, and about 20% from schools having fewer than fifteen students in French II. The corresponding per cents for French III are 11%, 0.8%, and 10%. The average scores in both French II and III of all three of these groups of schools are below the state-wide average, particularly of the senior schools and academies. Of the 20,000-odd students taking either the French II or French III Regents examinations in June, 1925, nearly 8000, or nearly 40%, had their French instruction in academies, senior schools, or schools having fewer than fifteen students in French II or French III. More than a fifth of the French II and III students in New York State are in schools in which the teacher of French has fewer than fifteen students in the second- or in the third-year class in this language.

This discussion of Table 21 applies with only small modifications to Tables 22, 23, and 24, which give data on the Spanish, German, and physics examinations paralleling those which Table 21 gives for the French examinations. The reader will readily understand the slight differences between the tables. They will therefore be presented without comment, beyond calling the attention of the reader to the fact that the Department ratings on the old-type examinations for Spanish II and III seem much less accurate, in terms of agreement with new-type results,

than the school ratings on these examinations. This is strong confirmation of the suggestions already made (pages 133–136) concerning the Regents reviewing system.

TABLE 22

INTERCORRELATIONS BETWEEN NEW-TYPE SCORES, SCHOOL RATINGS AND REVIEW RATINGS ON OLD-TYPE EXAMINATIONS IN SPANISH FOR INDIVIDUAL SCHOOLS AND CLASSES OF SCHOOLS, SHOWING VARIATIONS IN ACCURACY OF SCORING IN DIFFERENT SCHOOLS, AND SHOWING THAT SOME SCHOOLS THAT MOST NEEDED REVIEWING ESCAPED IT. THE VARIATIONS IN CORRELATIONS ARE PARTLY ACCOUNTED FOR BY DIFFERENCES IN VARIABILITIES OF INDIVIDUAL SCHOOL GROUPS. NT MEANS NEW-TYPE, S MEANS SCHOOL RATINGS, AND R REVIEW RATINGS.

School No.	Correlations between			Sigmas			Means			
	NT S	NT R	S R	NT	S	R	NT	S	R	N
Spanish II										
337	.847	.779	.936	32.08	9.21	8.52	148.95	73.3	71.12	109
466	.734	.705	.796	28.0	8.43	7.96	162.29	80.59	72.96	85
804	.613	.538	.845	26.75	4.3	3.4	130.0	66.7	65.55	40
806	.756	.721	.851	28.63	7.7	7.76	174.25	75.5	74.03	213
819	.844	.857	.986	24.92	8.2	8.32	156.8	72.53	72.05	38
862	.788	.799	.998	27.8	8.44	8.5	170.3	75.74	75.7	257
917	.479	.04	.089	22.5	2.84	.42	128.4	65.6	64.13	47
919	.613	.664	.702	30.7	6.04	5.88	131.5	68.4	69.34	89
Average 40 +	.709	.633	.775	(28.0)						
30–40	.663	.620	.902	28.56	7.92	7.85	159.48	74.51	73.69	268
15–30	.738	.720	.954	30.4	7.76	7.54	155.08	72.48	71.48	658
1–15	.651	.674	.811	28.5	7.22	7.25	154.3	72.61	71.02	609
Academies	.687	.682	.997	30.58	7.67	7.00	153.8	71.22	69.96	284
Total Sp. II	.699	.551		30.87	7.97	7.99	164.35	74.08	72.91	5340
Spanish III										
696	.811	.510	.947	23.8	8.12	8.39	174.87	72.93	70.53	30
793	.676	.668	.976	22.06	6.82	7.16	198.2	72.52	72.94	100
799	.609	.641	.822	18.91	6.8	8.07	202.8	72.41	76.89	47
800	.618	.596	.880	23.62	7.22	6.59	180.58	72.71	70.15	188
806	.731	.746	.938	27.24	8.62	8.54	207.46	79.42	79.45	77
810	.610	.651	.806	18.28	6.80	6.07	197.54	78.31	74.00	59
811	.748	.737	.998	20.66	8.2	7.93	200.43	73.19	73.05	160
892	.672	.668	.978	22.79	8.13	8.14	195.11	76.97	76.86	184
901	.713	.658	.958	19.86	6.6	6.83	189.25	73.69	73.22	107
971	.595	.581	.931	16.94	7.16	7.00	205.34	74.38	74.97	58
984	.509	.468	.934	25.52	6.74	6.61	213.95	77.25	78.00	48
Average 30 +	.663	.629	.924	(22.06)						
15–30	.697	.675	.975	24.25	7.27	7.48	185.06	72.90	73.36	332
1–15	.653	.682	.870	27.00	7.92	7.93	188.03	74.73	72.04	312
Academies	.680	.654	.840	28.45	10.11	8.96	188.0	76.08	72.36	170
Total Sp. III	.670	.654		25.2	8.18	8.09	192.7	74.65	73.88	2132

TABLE 22 (*Continued*)

Spanish IV

Total Sp. IV	.696	.713		20.5	8.88	9.64	222.17	80.27	79.19	226

TABLE 23

INTERCORRELATIONS BETWEEN NEW-TYPE SCORES, SCHOOL RATINGS AND REVIEW RATINGS ON OLD-TYPE EXAMINATIONS IN GERMAN FOR INDIVIDUAL SCHOOLS AND CLASSES OF SCHOOLS, SHOWING VARIATIONS IN ACCURACY OF SCORING IN DIFFERENT SCHOOLS, AND SHOWING THAT SOME SCHOOLS THAT MOST NEEDED REVIEWING ESCAPED IT. THE VARIATIONS IN CORRELATIONS ARE PARTLY ACCOUNTED FOR BY DIFFERENCES IN VARIABILITIES OF INDIVIDUAL SCHOOL GROUPS. NT STANDS FOR NEW-TYPE, S FOR SCHOOL RATINGS, AND R FOR REVIEW RATINGS.

German II

School No.	Correlations between			Sigmas			Means			
	NT S	NT R	S R	NT	S	R	NT	S	R	N
696	.747	.747	1.000	32.18	8.48	8.48	135.37	73.85	73.85	54
793	.675	.679	.989	19.05	7.39	7.40	186.82	76.52	76.55	66
830	.777	.771	.985	19.3	6.32	6.45	198.33	81.89	82.0	36
862	.609	.646	.994	23.8	6.38	6.18	200.36	85.71	85.71	56
884	.568	.581	.986	32.0	11.66	11.26	153.23	78.82	78.21	34
892	.671	.661	.982	18.75	6.82	6.66	207.21	84.71	84.79	68
896	.657	.624	.986	27.45	7.9	8.19	193.38	81.24	81.46	37
901	.782	.794	.998	34.8	8.22	8.2	147.86	77.31	77.03	70
954	.830	.815	.992	24.59	6.71	6.4	182.44	83.16	83.35	43
962	.469	.445	.957	23.53	6.25	6.07	192.43	81.66	82.06	35
971	.488	.420	.963	18.68	5.34	5.48	194.0	82.6	84.12	50
984	.781	.757	.950	19.485	5.37	4.527	219.5	85.6	86.05	40
1039	.638	.679	.976	26.72	11.2	10.66	154.43	78.35	77.80	69
Average 30 +	.669	.663	.981	(24.0)						
15–30	.708	.689	.981	35.49	9.54	9.51	180.91	77.77	78.22	303
1–15	.629	.691	.910	39.55	7.45	6.53	156.5	75.86	75.59	324
Academies	.655	.753	.910	46.6	9.62	9.73	155.12	76.29	75.42	163
Total Ger. II	.669	.719	.951	39.09	8.84	9.02	172.1	78.498	78.42	1482

German III

Total Ger. III	.711	.756		33.56	9.2	9.29	202.9	77.21	77.15	667

German IV

Total Ger. IV	.640	.723		25.4	8.52	8.38	228.07	79.65	79.35	127

TABLE 24

INTERCORRELATIONS BETWEEN NEW-TYPE SCORES, SCHOOL RATINGS AND REVIEW RATINGS ON OLD-TYPE EXAMINATION IN PHYSICS FOR INDIVIDUAL SCHOOLS AND CLASSES OF SCHOOLS, SHOWING VARIATIONS IN ACCURACY OF SCORING IN DIFFERENT SCHOOLS, AND SHOWING THAT SOME SCHOOLS THAT MOST NEEDED REVIEWING ESCAPED IT. THE VARIATIONS IN CORRELATIONS ARE PARTLY ACCOUNTED FOR BY DIFFERENCES IN VARIABILITIES OF INDIVIDUAL SCHOOL GROUPS.

Physics										
School No.	Correlations between			Sigmas			Means			
	NT S	NT R	S R	NT	S	R	NT	S	R	N
337	.474	.415	.849	19.45	10.27	9.04	92.64	80.83	80.65	147
369	.318	.290	.493	17.65	7.78	4.28	126.3	67.94	65.91	64
696	.507	.475	.933	20.65	10.31	9.5	81.74	81.13	83.14	152
800	.583	.561	.955	22.5	10.86	10.65	87.03	81.22	81.27	411
819	.632	.650	.918	21.6	11.9	10.14	90.59	79.09	76.64	81
804	.501	.506	.906	20.45	9.16	9.12	104.46	71.5	71.78	102
862	.342	.271	.840	18.6	10.02	9.34	82.69	82.54	82.19	129
863	.654	.556	.899	20.05	10.83	8.9	95.45	83.18	81.41	61
884	.636	.631	.930	24.1	11.76	11.5	89.05	75.5	76.05	140
918	.563	.594	.900	21.85	9.74	9.88	93.42	75.4	74.39	164
942	.517	.515	.978	20.55	11.21	11.00	87.34	81.43	81.56	123
971	.524	.561	.926	19.81	10.89	10.08	85.01	82.88	83.02	338
Averages	.521	.502	.877	(20.55)						1912
Total Physics	.556	.582	.944	22.9	10.7	10.35	97.59	77.32	76.66	14081

School differences with respect to achievement in vocabulary, reading and grammar. — In the preceding discussion attention was called to the fact that teaching conditions in small and private schools are considerably less favorable than in large schools because of the greater heterogeneity of the classes in the former. The classification of students for instructional purposes is better in the large than in the small schools, but even among the large schools there are notable variations in kinds and amounts of modern language achievement which have considerable significance both for the theory and practice of measurement and for the structure of the curriculum in the modern languages. Table 25 summarizes some of these variations for second- and third-year French classes in the large schools mentioned in connection with Table 21.

The general conclusion from Table 25 is that even in the largest modern language departments instruction is not adapted to individual needs of students, and that different schools emphasize different aspects of the language, so that achievement is very uneven both for individual students and for individual schools.

Homogeneity of classes. — Columns (5), (6), and (7) in Table 25 show large variations in degree of homogeneity of the classes in different schools

TABLE 25

INTERCORRELATIONS OF THE VOCABULARY, READING AND GRAMMAR PARTS OF THE NEW-TYPE FRENCH EXAMINATION, BASED ON RESULTS FROM INDIVIDUAL SCHOOLS, WITH MEANS AND SIGMAS, SHOWING "UNEVENNESS" OF ACHIEVEMENT OF INDIVIDUALS AND OF SCHOOLS IN THESE THREE ELEMENTS OF MODERN LANGUAGE ACHIEVEMENT.

School No.	Correlations between Parts			Sigmas			Means			N
	I and II	I and III	II and III	I	II	III	I	II	III	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
French II										
337	.528	.552	.492	8.50	10.40	10.48	57.89	45.25	50.44	102
369	.772	.733	.777	13.45	15.35	13.60	38.97	22.67	18.88	58
591	.607	.686	.637	8.30	9.65	12.18	57.08	43.35	45.38	59
791	.551	.553	.483	7.85	10.90	10.35	68.57	53.57	58.71	70
806	.546	.610	.631	10.4	9.3	10.1	65.51	48.48	56.69	178
819	.597	.488	.475	8.28	9.25	10.23	59.62	42.67	45.27	144
862	.522	.597	.491	8.35	8.73	8.88	66.13	50.39	59.21	335
884	.542	.575	.574	9.95	9.3	12.15	57.61	46.25	52.98	143
919	.633	.598	.725	10.7	11.75	11.55	58.39	42.22	51.95	129
696	.638	.654	.617	7.22	9.8	10.75	62.72	50.45	54.78	132
942	.677	.597	.623	8.53	10.5	10.83	63.79	48.38	53.14	62
971	.573	.667	.620	7.92	8.77	10.5	70.67	52.42	59.66	241
Total Fr. II	.658	.714	.699	11.05	11.85	13.5	57.61	44.64	48.47	13486
French III										
337	.432	.623	.547	7.02	7.02	8.91	76.94	58.55	65.13	61
696	.647	.682	.517	5.32	8.91	9.69	76.06	57.88	62.43	68
701	.575	.606	.591	6.26	8.51	10.53	76.31	59.5	62.31	48
791	.431	.523	.509	6.63	6.96	8.97	75.63	58.17	62.03	45
806	.702	.553	.551	11.01	7.86	12.27	74.5	55.72	61.23	66
819	.607	.561	.529	7.56	8.67	10.68	72.66	55.22	61.5	43
862	.457	.465	.425	7.14	7.31	6.54	76.87	56.03	68.05	235
881	.437	.612	.572	9.24	10.38	10.71	68.54	52.61	58.87	46
883	.697	.534	.583	9.00	7.83	9.57	68.43	52.85	55.97	43
884	.563	.562	.494	8.43	7.44	9.12	74.14	55.95	64.82	56
919	.618	.657	.549	9.02	8.81	9.86	71.66	52.45	59.39	57
971	.469	.504	.346	7.56	6.60	8.46	79.81	59.16	66.84	115
984	.489	.642	.506	8.21	8.19	10.89	73.78	56.21	68.18	263
Total Fr. III	.624	.652	.609	10.20	9.65	11.55	70.37	53.75	60.26	6741
French IV										
Total Fr. IV	.550	.698	.531	10.26	8.18	11.91	79.16	59.68	67.50	489

with respect to each of the three aspects of achievement measured by the new-type examinations. The French II class in school 696 is very homogeneous with respect to vocabulary, and fairly so with respect to reading and grammar; by contrast, the French II class of school 806 is more

homogeneous with respect to reading and grammar than 696, but very much more heterogeneous with respect to vocabulary than 696. There are several variations of this sort in Table 25 which the reader may find for himself. Space limitations force us to confine ourselves to specific mention of only one, and to point out that the teaching problem in school 696 is very different from that in 806. Half the teaching battle is won when we have found just what the student already knows and therefore just what he still needs to learn; the battle is as good as lost when we "shoot in the dark" by teaching without first ascertaining just what ought to be taught to a given student or group of students. The new-type objective tests will avail much in this matter of defining immediate goals and in directing the efforts of teachers under varying classroom conditions. It is this capacity of the new-type tests to give constructive information that raises them above the level of mere "flunking" and punitive devices, and recommends them to wide use not only as final examinations but as classroom instruments of indispensable pedagogic values.

Unevenness of achievement.— Columns (8), (9), and (10) in Table 25 show that some schools are of high average achievement in one respect, e.g. active grammar, and low in vocabulary or reading; and others are high in vocabulary but low in reading and grammar. This is undoubtedly mainly due to differing emphases on the part of the teachers in different schools. Some teachers believe in extensive reading, and tend to neglect the other things; other teachers have great faith in active grammar, and exercise their students so intensively in it that little time or energy is left for vocabulary, reading, oral-aural practice, etc. Thus, without knowing or desiring it, teachers sacrifice fully-rounded progress of their pupils to their personal predilections and individual theories of the proper study of languages. The use of objective and standardized tests which represent each of the various aspects of modern language achievement will save both teachers and pupils from such errors. It has obviously been impossible to secure such constructive guidance from the old-type subjective examinations. Their subjectivity alone is sufficient to prevent them from performing such service. There are some teachers who believe they can tell from a student's translations of two or three brief passages how much grammar, vocabulary, reading ability, etc., he has; but this is obviously unwarranted faith.

Intercorrelations of new-type Parts for individual schools. — The intercorrelations of the Parts shown in columns (2), (3), and (4) are fairly constant for all the schools represented in Table 25. This fact confirms the indications of earlier tables that the vocabulary, reading, and grammar aspects are definitely interdependent, yet specific and independent enough to require separate measurements. There appears to be no departure

from this rule. In no school would a test of vocabulary alone, reading alone, or grammar alone, give an adequate measure of total modern language achievement. From this we may infer that an examination which consists mainly of written translation exercises is not broad enough to afford adequate measures of total language achievement, for if vocabulary and silent reading are independent enough to require separate measurement, surely so distinct a skill and so questionable a discipline as turning one language into another cannot be considered as representative of total

TABLE 26

INTERCORRELATIONS OF THE VOCABULARY, READING AND GRAMMAR PARTS OF THE NEW-TYPE SPANISH EXAMINATION, BASED ON RESULTS FROM INDIVIDUAL SCHOOLS, WITH MEANS AND SIGMAS, SHOWING "UNEVENNESS" OF ACHIEVEMENT OF INDIVIDUALS AND OF SCHOOLS IN THESE THREE ELEMENTS OF MODERN LANGUAGE ACHIEVEMENT.

School No.	Correlations between Parts			Sigmas			Means			N
	I and II	I and III	II and III	I	II	III	I	II	III	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Spanish II										
337	.705	.768	.634	12.39	10.7	12.45	64.13	47.6	34.85	109
466	.628	.739	.622	10.4	9.6	10.95	72.25	51.7	38.9	85
804	.468	.557	.474	11.1	10.55	10.86	56.6	40.55	34.1	40
806	.517	.684	.515	10.63	8.79	13.83	74.57	50.38	50.59	213
819	.702	.855	.565	8.6	12.21	12.4	65.7	51.03	40.8	38
862	.501	.631	.575	8.67	8.61	15.2	73.15	51.67	46.4	257
917	.561	.497	.425	10.7	8.85	6.81	57.2	44.75	25.1	47
919	.616	.703	.658	10.3	11.16	12.8	56.9	42.7	33.1	89
Total Sp. II	.629	.702	.578	11.43	9.69	14.16	70.11	51.68	43.12	5340
Spanish III										
696	.616	.631	.596	8.00	6.65	12.33	75.4	54.7	40.3	30
793	.417	.669	.521	6.88	6.89	11.77	83.36	59.5	56.06	100
799	.397	.771	.381	6.88	5.0	11.2	83.61	59.36	57.96	47
800	.573	.638	.515	9.57	7.97	10.35	74.21	54.48	49.86	188
806	.669	.595	.565	9.29	8.15	12.7	84.39	57.42	62.98	77
810	.202	.340	.607	7.87	6.48	9.14	81.15	58.99	55.01	59
811	.421	.660	.425	6.6	6.6	11.2	83.63	54.91	56.46	160
892	.410	.630	.490	7.37	7.89	7.71	83.52	56.52	58.5	184
901	.421	.584	.411	7.25	7.08	10.55	81.03	57.07	51.45	107
971	.361	.386	.458	4.81	6.00	9.96	83.14	60.5	63.03	58
984	.624	.739	.622	6.55	7.02	11.67	86.19	60.83	66.02	48
Total Sp. III	.522	.605	.486	8.79	7.83	12.78	80.9	57.66	53.95	2132
Spanish IV										
Total Sp. IV	.378	.573	.417	6.57	6.81	12.15	89.69	63.82	69.77	226

modern language achievement in the high school stages. In a final examination which aims to measure total achievement and to give the teacher information that may be used constructively in the classroom, tests of vocabulary and silent reading ability are just as necessary as tests of grammar, translation, and oral and aural skills.

TABLE 27

INTERCORRELATIONS OF THE VOCABULARY, READING AND GRAMMAR PARTS OF THE NEW-TYPE GERMAN EXAMINATION, BASED ON RESULTS FROM INDIVIDUAL SCHOOLS, WITH MEANS AND SIGMAS, SHOWING "UNEVENNESS" OF ACHIEVEMENT OF INDIVIDUALS AND OF SCHOOLS IN THESE THREE ELEMENTS OF MODERN LANGUAGE ACHIEVEMENT.

School No.	Correlations between Parts			Sigmas			Means			N
	I and II	I and III	II and III	I	II	III	I	II	III	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
German II										
696	.759	.787	.702	11.15	9.975	12.9	65.89	41.89	30.45	54
793	.479	.362	.345	7.08	7.005	11.25	82.77	53.23	51.68	66
830	.607	.505	.227	7.16	6.95	10.13	84.0	54.67	60.58	36
862	.675	.435	.409	7.51	7.32	12.89	87.21	55.82	59.04	56
884	.806	.756	.561	10.3	10.66	13.73	73.06	46.99	37.59	34
892	.442	.356	.319	6.52	6.21	11.92	88.5	57.13	65.69	68
896	.586	.649	.619	9.48	8.05	13.8	85.26	55.01	53.88	37
901	.622	.611	.469	14.7	11.8	14.88	70.2	45.0	36.0	70
954	.607	.482	.330	9.59	9.69	10.9	81.52	52.99	51.38	43
962	.607	.656	.418	8.37	9.75	8.89	81.13	54.13	61.41	35
971	.593	.344	.327	7.55	6.47	10.3	85.44	55.08	57.9	50
984	.523	.429	.483	6.596	7.999	9.86	89.625	57.15	73.5	40
1039	.613	.499	.509	10.0	10.3	12.54	68.43	46.61	38.67	69
Total II	.738	.685	.583	14.1	8.21	17.7	76.44	48.51	46.86	1482
German III										
Total III	.741	.730	.632	9.77	9.45	9.46	88.28	57.2	57.11	667
German IV										
Total IV	.443	.487	.468	6.64	7.03	15.95	94.39	63.54	68.95	127

Tables 26 and 27, relating to Spanish and German results, are similar to Table 25 and confirm its indications. Indeed, the variations in degree of heterogeneity of classes with respect to vocabulary, reading, and grammar are greater in Spanish and in German than in French classes. For example, the Spanish II class in school 917 is only half as variable with respect to grammar as most of the other schools; this is also true of the

Spanish III class in school 892, and to a lesser degree of the German II class in school 962. Several of the Spanish III classes represented in Table 26 are about twice as heterogeneous with respect to vocabulary as the Spanish III class in school 971. The increase in homogeneity from second- to fourth-year classes is much greater in Spanish and German than in French, although the numbers of students in the fourth-year classes in Spanish and German are too small to afford very reliable indications. The general conclusion from Tables 25–27 is that the new-type tests have uncovered variations in teaching conditions and in teaching problems in different schools which the old-type tests could not reveal; that, in addition to giving more reliable and more complete measures of total language achievement than the old-type, the new-type tests give more specific information which can be used in a constructive manner in the classroom in individualizing instruction and in directing the efforts of both teachers and students into the most profitable channels. The new-type tests will thus perform as great a service when given at the beginning and during the school year as at the end.

IV

OVERLAPPING OF CLASSES AND VARIABLE SCHOOL STANDARDS NOT CORRECTED BY REGENTS EXAMINATIONS

Distributions of New-Type Examination Scores by Year-Classes

For the convenience of investigators and others who may use the new-type Regents examinations, or one of their equivalent forms, we shall now present without comment a series of eight tables showing various distributions of scores on the new-type examinations.

Distributions of new-type total scores. — Tables 28–31, inclusive, show distributions of total scores on the new-type examinations in French, Spanish, German, and physics, (*a*) of all students in each class taking Regents examinations in June, 1925, (*b*) of all students in each class passed by the old-type part of the Regents examinations, and (*c*) of all students failed by the old-type part of the Regents examinations.

Some of the facts about the modern language situation in New York State which these tables reveal will be elaborated and set forth graphically in later sections of this report; but the tables themselves are well worth whatever time the reader may spend in studying them.

No attempt can be made within the space limits at our disposal to notice all the interesting features of these tables. Attention, however, may be called to one series of comparisons which the reader may make. In Table 28 the sixth line from the bottom shows the approximate per cents of students passed and failed in each year-class by the old-type part of the Regents examinations. Why is the proportion of failures in French IV greater than in French III? The proportions of students failed by the old-type Regents examinations in second-, third-, and fourth-year French are, respectively, 24%, 15% and 20%. A mortality rate of one in five for fourth-year students, many of whom have attended "five recitations a week" for five or more years, is quite serious. One would expect that after students have been allowed to stay in modern language work for three or more years the proportions of failures would be progressively smaller. Table 29 fulfils this expectation by showing that the proportions of failures in Spanish steadily decrease from second- to fourth-year classes—20%, 15%, and 7%. In fulfilling this expectation Spanish stands alone, for in Table 30 we find the trend of failing rates exactly reversed, the proportions of failures for second-, third-, and fourth-year German classes being 11%, 13%, and 13%, respectively. These figures constitute a serious indictment against the educational guidance which is accorded to modern language students in New York State.

TABLE 28

DISTRIBUTIONS OF TOTAL SCORES ON NEW-TYPE REGENTS EXAMINATION IN FRENCH, JUNE, 1925, OF ALL STUDENTS, OF "PASSED" AND OF "FAILED" STUDENTS BY YEAR-CLASSES, SHOWING MEANS AND SIGMAS, AND APPROXIMATE QUARTILE AND MEDIAN SCORES. THE FIGURES IN PARENTHESES SHOW APPROXIMATE PER CENTS OF EACH CLASS PASSED AND FAILED BY THE OLD-TYPE REGENTS EXAMINATION.

French

NEW-TYPE EXAM.	FRENCH II			FRENCH III			FRENCH IV		
TOTAL SCORES	Total Population	Passed Old-Type Exam.	Failed Old-Type Exam.	Total Population	Passed Old-Type Exam.	Failed Old-Type Exam.	Total Population	Passed Old-Type Exam.	Failed Old-Type Exam.
10–19	1		1						
20–29	4		4	1		1			
30	10		10						
40	19		19	3		3			
50	31		31	6		6			
60	76	3	73	7		7			
70	111	6	105	5		5			
80	230	31	199	8		8			
90	352	72	280	20		20	1	1	
100	555	178	377	29	4	25	1	1	
110	788	326	462	51	9	42	1		1
120	1190	669	521	100	27	73	1		1
130	1462	1017	445	183	75	108	2		2
140	1577	1275	302	295	148	147	7	1	6
150	1691	1492	199	450	293	157	14	5	9
160	1613	1514	99	752	590	162	31	18	13
170	1291	1249	42	921	789	132	38	22	16
180	1021	1003	18	1069	988	81	37	23	14
190	735	728	7	1000	970	30	54	39	15
200	424	424		826	815	11	75	69	6
210	197	197		533	530	3	72	61	11
220	79	78	1	301	300	1	64	60	4
230	23	23		139	137	2	49	48	1
240	6	6		35	35		36	36	
250–259				7	7		6	6	
Total	13486	10291	3195	6741	5717	1024	489	390	99
Per cents		(76%)	(24%)		(85%)	(15%)		(80%)	(20%)
Means	150.4			183.4			205.8		
Sigmas	31.6			27.4			27.4		
L. Q.	130	102	102	167	173	136	187	196	165
Mdn.	152	160	120	185	189	154	208	213	182
U. Q.	173	179	137	203	214	170	225	229	197

TABLE 29

DISTRIBUTIONS OF TOTAL SCORES ON NEW-TYPE REGENTS EXAMINATION IN SPANISH, JUNE, 1925, OF ALL STUDENTS, OF "PASSED" AND OF "FAILED" STUDENTS BY YEAR-CLASSES, SHOWING MEANS AND SIGMAS, AND APPROXIMATE QUARTILE AND MEDIAN SCORES. THE FIGURES IN PARENTHESES SHOW APPROXIMATE PER CENTS OF EACH CLASS PASSED AND FAILED BY THE OLD-TYPE REGENTS EXAMINATION.

Spanish

NEW-TYPE EXAM.	SPANISH II			SPANISH III			SPANISH IV		
TOTAL SCORES	Total Population	Passed Old-Type Exam.	Failed Old-Type Exam.	Total Population	Passed Old-Type Exam.	Failed Old-Type Exam.	Total Population	Passed Old-Type Exam.	Failed Old-Type Exam.
40–49	1	1							
50–59	5		5						
60	9		9						
70	18	1	17	1		1			
80	25	2	23						
90	55	4	51	1		1			
100	122	19	103	4		4			
110	181	56	125	9	2	7			
120	319	137	182	8	1	7			
130	401	241	160	18	8	10			
140	609	451	158	45	14	31			
150	649	541	108	111	62	49			
160	688	622	66	177	116	61	2	1	1
170	593	563	30	248	195	53	5	4	1
180	584	576	8	329	285	44	10	8	2
190	453	445	8	364	336	28	17	13	4
200	286	284	2	309	299	10	26	24	2
210	195	194	1	223	218	5	37	34	3
220	93	91	2	130	130		38	36	2
230	38	38		106	105	1	43	42	1
240	15	15		37	36	1	34	34	
250	1	1		10	10		12	12	
260–269				2	2		2	2	
Total	5340	4282	1058	2132	1819	313	226	210	16
Per cents		(80%)	(20%)		(85%)	(15%)		(93%)	(7%)
Means	164.4			192.7			222.2		
Sigmas	30.9			25.2			20.5		
L. Q.	143	152	115	176	182	153	209	210	190
Mdn.	163	171	131	193	197	166	224	226	200
U. Q.	185	190	148	208	212	183	238	238	216

TABLE 30

DISTRIBUTIONS OF TOTAL SCORES ON NEW-TYPE REGENTS EXAMINATION IN GERMAN, JUNE, 1925, OF ALL STUDENTS, OF "PASSED" AND OF "FAILED" STUDENTS BY YEAR-CLASSES, SHOWING MEANS AND SIGMAS, AND APPROXIMATE QUARTILE AND MEDIAN SCORES. THE FIGURES IN PARENTHESES SHOW APPROXIMATE PER CENTS OF EACH CLASS PASSED AND FAILED BY THE OLD-TYPE REGENTS EXAMINATION.

German

NEW-TYPE EXAM.	GERMAN II			GERMAN III			GERMAN IV		
TOTAL SCORE	Total Population	Passed Old-Type Exam.	Failed Old-Type Exam.	Total Population	Passed Old-Type Exam.	Failed Old-Type Exam.	Total Population	Passed Old-Type Exam.	Failed Old-Type Exam.
50–59	4	1	3						
60–69	6		6						
70	10	2	8						
80	18	8	10						
90	36	19	17						
100	42	22	20	6	2	4			
110	47	31	16	7	4	3			
120	77	63	14	6	3	3			
130	69	57	12	7	1	6			
140	91	81	10	13	3	10	2		2
150	127	114	13	37	21	16	3	1	2
160	137	128	9	33	19	14	1		1
170	126	123	3	54	45	9	4	1	3
180	154	151	3	55	51	4	6	2	4
190	151	149	2	79	67	12	11	10	1
200	141	141		74	71	3	10	8	2
210	93	93		72	70	2	28	27	1
220	75	75		62	62		14	13	1
230	47	47		67	67		24	24	
240	23	23		52	52		15	15	
250	6	6		35	35		8	8	
260	2	2		7	7		1	1	
270–279				1	1				
Total	1482	1336	146	667	581	86	127	110	17
Per cents		(89%)	(11%)		(87%)	(13%)		(87%)	(13%)
Means	172.1			202.9			228.1		
Sigmas	39.1			33.6			25.4		
L. Q.	147	154	96	180	189	145	205	212	160
Mdn.	176	181	116	215	210	160	219	225	180
U. Q.	201	203	141	230	232	180	237	238	195

TABLE 31

DISTRIBUTIONS OF TOTAL SCORES ON NEW-TYPE REGENTS EXAMINATION IN PHYSICS, JUNE, 1925, OF ALL STUDENTS, OF "PASSED" AND OF "FAILED" STUDENTS BY YEAR-CLASSES, SHOWING MEANS AND SIGMAS, AND APPROXIMATE PER CENTS OF EACH CLASS PASSED AND FAILED BY THE OLD-TYPE REGENTS EXAMINATION.

Physics

NEW-TYPE EXAM. TOTAL SCORES	TOTAL POPULATION	PASSED OLD-TYPE EXAM.	FAILED OLD-TYPE EXAM.
125–129	3	3	
120–124	7	7	
115–119	22	22	
110–114	43	43	
105–109	53	53	
100–104	110	109	1
95–99	142	141	1
90–94	241	239	2
85–89	278	276	2
80–84	431	426	5
75–79	465	456	9
70–74	654	641	13
65–69	713	689	24
60–64	944	895	49
55–59	968	898	70
50–54	1136	1043	93
45–49	1222	1104	118
40–44	1240	1036	204
35–39	1103	887	216
30–34	1085	826	259
25–29	836	577	259
20–24	818	519	299
15–19	563	348	215
10–14	460	252	208
5–9	274	122	152
0–4	270	105	165
Numbers .	14,081	11,717	2,364
Per cent . .		85%	15%
Means . .	97.12	93.29	115.98
Sigmas . .	23.0	22.15	16.77

Distributions of scores on each Part, and on Parts I and II combined, of new-type tests. — Tables 32–34, inclusive, show distributions of scores on each Part of the new-type modern language tests by total year-classes. Table 35 shows distributions of combined Part I and Part II scores, on the new-type modern language examinations for random samplings of each year-class. These last four tables are included because of their general usefulness to investigators, and for the convenience of teachers

who may administer only one or two Parts of the new-type tests, or who wish to compare the achievements of their own classes in each Part with state-wide norms for each Part.

TABLE 32

DISTRIBUTIONS OF SCORES ON EACH PART OF THE NEW-TYPE REGENTS EXAMINATION OF JUNE, 1925, BY YEAR-CLASSES

French

CLASSES	PART I VOCABULARY			PART II TRUE-FALSE			PART III COMPLETION		
	II	III	IV	II	III	IV	II	III	IV
Scores									
-1–0–1				31	6		2	1	
2–3–4				25	2		6	3	
5–6–7				43			29	2	
9	1	1		54	1		40	6	
12	5	1		80	9		52	7	
15	4			121	6	1	107	4	
18	4	2		144	15		136	5	
21	16	1		214	13		195	12	
24	26	1		381	25	1	306	19	
27	35	5		433	38	1	362	16	4
30	86	3		675	52		564	37	1
33	132	7		793	92	2	618	49	2
36	241	8		1071	166	2	788	71	1
39	385	14		1051	186	5	974	116	3
42	458	29		1054	369	12	1079	204	4
45	856	44	1	1497	417	9	1141	244	6
48	981	92	1	1395	720	21	1139	377	11
51	1291	122	4	1197	653	27	1169	418	25
54	1383	211	7	1097	1005	60	1057	579	27
57	1452	304	7	1056	801	50	1018	703	35
60	1390	455	11	441	909	82	750	733	36
63	1282	576	18	370	594	71	678	764	33
66	1085	724	22	173	436	87	466	659	50
69	795	802	33	78	155	36	326	593	43
72	680	780	41	27	50	22	279	444	50
75	358	736	39	5	21		91	302	45
78	281	632	50				71	172	35
81	116	530	70				30	115	38
84	73	325	43				8	53	23
87	48	201	47				5	26	15
90	12	93	48					5	1
93	6	36	37					2	1
95–96–97	3	5	10						
98–99–100		1							
Numbers	13486	6741	489	13486	6741	489	13486	6741	489
Means	57.61	70.37	79.16	44.64	53.75	59.68	48.47	60.26	67.50
Sigmas	11.05	10.20	10.26	11.85	9.65	8.18	13.5	11.55	11.91

TABLE 33

DISTRIBUTIONS OF SCORES ON EACH PART OF THE NEW-TYPE REGENTS EXAMINATION OF JUNE, 1925, BY YEAR-CLASSES

Spanish

CLASSES	PART I			PART II			PART III		
	II	III	IV	II	III	IV	II	III	IV
Scores									
2–3–4				2			3		
5–6–7				1			8		
9				2			18	1	
12				4			41	1	
15	1			6			82	6	
18				8			123	3	
21	2			16	2		140	7	
24	2			27	2		223	20	
27	5			35			282	14	
30	6			57	3		323	33	
33	8			129	10		426	47	
36	13			167	14		445	77	1
39	13			277	24		398	98	2
42	33	1		395	60		458	156	2
45	63	3		472	79	5	596	158	4
48	123	4		580	157	5	388	189	5
51	164	7		592	180	4	350	233	9
54	194	7		697	289	13	262	192	10
57	281	10		589	318	23	246	172	16
60	349	28		591	337	30	213	181	12
63	440	44	1	293	255	37	156	130	13
66	466	66		258	253	51	133	118	18
69	492	90	3	99	108	29	94	79	25
72	587	150	1	41	38	20	47	75	16
75	542	238	6	2	3	9	34	63	21
78	495	249	6				27	28	22
81	394	315	16				16	22	20
84	297	271	18				5	19	10
87	186	278	35				3	8	11
90	97	184	48					2	8
93	52	123	53						1
95–96—97	31	49	22						
98–99–100	4	15	17						
Numbers	5340	2132	226	5340	2132	226	5340	2132	226
Means	70.11	80.9	89.69	51.68	57.66	63.82	43.12	53.95	69.77
Sigmas	11.43	8.79	6.57	9.69	7.83	6.81	14.16	12.78	12.15

TABLE 34

DISTRIBUTIONS OF SCORES ON EACH PART OF THE NEW-TYPE REGENTS EXAMINATION OF JUNE, 1925, BY YEAR-CLASSES

German

CLASSES	PART I			PART II			PART III		
	II	III	IV	II	III	IV	II	III	IV
Scores									
−1−0−1				1					
2–3–4				1			2		
5–6–7				1			7		
9				5			9	1	
12				6			21	4	
15				9			32		
18				9	1		34	5	
21	1			12			43	5	1
24				31	1		59	7	
27	1			31	3		54	11	1
30	2			46	3		68	19	1
33	1			50	4		72	20	1
36	2			67	14	1	74	29	
39	8			76	11	1	90	22	2
42	10	1		107	20	1	90	34	4
45	15			118	33	2	92	33	4
48	29			140	49		86	34	2
51	43	1		145	53	3	81	47	6
54	38	3		154	59	8	90	39	2
57	52	8		142	72	7	91	59	3
60	51	2		137	89	15	70	39	6
63	65	6	1	95	92	12	66	25	15
66	77	8	1	61	87	46	64	29	7
69	88	6		25	51	17	49	30	8
72	85	19		11	23	12	38	41	8
75	115	29	2	2	2	2	32	35	13
78	102	33	3				30	23	9
81	132	50	3				20	14	5
84	128	61	1				7	22	6
87	132	63	9				6	21	9
90	114	80	6				2	10	5
93	97	105	20				3	7	6
95–96–97	69	116	39					2	3
98–99–100	25	76	42						
Numbers	1482	667	127	1482	667	127	1482	667	127
Means	76.44	88.28	94.39	48.51	57.2	63.54	46.86	57.11	68.95
Sigmas	14.1	9.77	6.64	8.21	9.45	7.03	17.7	17.74	15.95

TABLE 35

DISTRIBUTIONS OF COMBINED SCORES ON THE VOCABULARY AND READING PARTS OF THE NEW-TYPE EXAMINATIONS IN FRENCH, SPANISH, AND GERMAN OF RANDOM SAMPLINGS OF STUDENTS WHO TOOK THE REGENTS EXAMINATIONS IN JUNE, 1925.

SCORES	French			Spanish			German		
	II	III	IV	II	III	IV	II	III	IV
15–19	1								
20–24	3								
25	1								
30	6			2					
35	6			1					
40	9			1			1		
45	6			1			3		
50	17			3			5		
55	10			6			4		
60	23			7			7		
65	26			8	1		14		
70	43	1	1	10	1		17		
75	55	4		20			20	2	
80	103	5	1	28	1		34	1	
85	123	11	1	41	2		45	4	
90	153	35	1	64	12		35	2	
95	169	49	5	88	14		62	5	1
100	171	77	5	122	10		54	8	
105	194	117	18	139	27		60	5	
110	188	183	13	162	65		79	10	
115	170	228	25	193	84		91	16	
120	146	247	30	235	137	5	96	24	2
125	137	248	33	211	182	6	95	45	1
130	109	226	39	213	236	2	94	46	1
135	68	208	47	162	283	9	123	45	2
140	39	144	53	140	290	25	112	46	2
145	18	89	53	97	282	28	125	60	11
150	13	60	48	52	225	41	76	68	3
155	5	22	44	13	48	52	57	76	17
160	1	6	24	8	77	34	47	74	21
165			9	1	29	14	26	59	17
170–174			1		2	10	7	16	5
Numbers	2013	1960	451	2028	2008	226	1389	612	83
Means	105.78	125.14	138.32	120.65	137.55	153.14	125.15	145.42	157.43
Sigmas	21.67	14.91	17.01	19.22	14.32	10.74	24.95	18.25	12.69

OVERLAPPING OF CLASSES

Means and sigmas of class distributions presented graphically. — Chart 16 is based on the data of Tables 28–31 and summarizes graphically the class-averages and variabilities of all French, Spanish, and German students who took the Regents examinations in June, 1925. The new-type tests are well adapted to the whole range of achievement in high school modern language work, and display the class-differences very well. The

differences between year-class averages are 33 and 23 score-points for French, 29 and 29 for Spanish, and 31 and 25 score-points for German classes. The sigma ranges show extensive overlapping of classes in all three languages.

Fallacy of customary assumption underlying use of separate examinations for year-classes. — The customary assumption regarding modern language

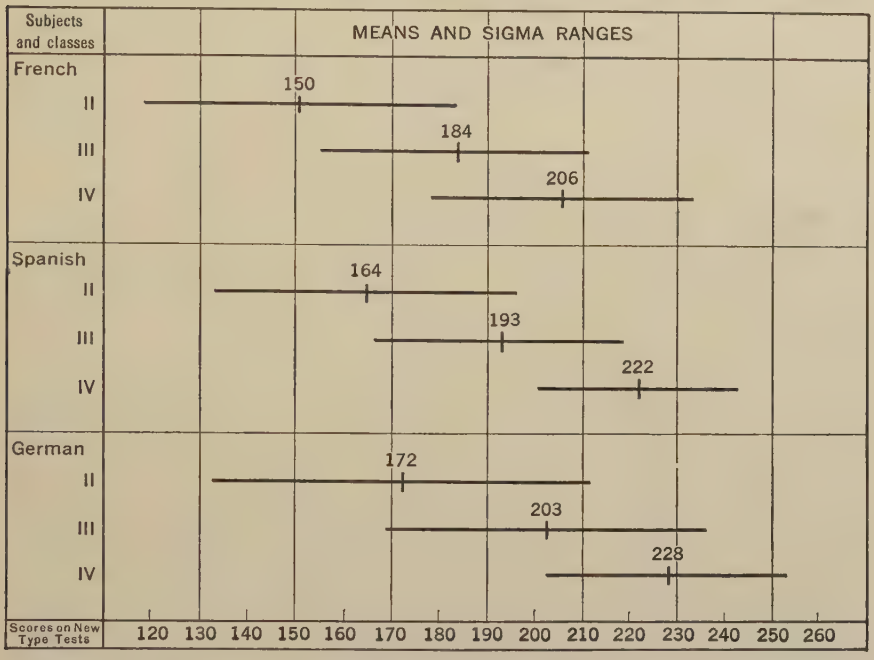
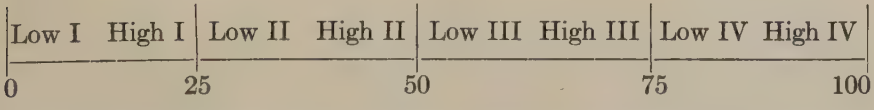


CHART 16. — Means and sigma-ranges of scores on new-type modern language examinations in French, Spanish, and German, of all students taking the Regents examinations in June, 1925. The sigma-range comprises about two-thirds of the scores in each class; the overlapping of classes is very apparent.

students who are in separate year-classes is that they are separate and distinct species, so that students in French III, for example, are regarded as *ipso facto* entirely above French II students and entirely below those in French IV. This assumption is illustrated graphically in the following figure.



The horizontal line represents the total range of achievement in high school modern language work, from the beginning of French I to the end of French IV. The lowest French II students are still above the highest

French I students, and the highest French II students are still below the lowest French III students. The examining problem is, therefore, one of merely differentiating among the best and poorest of each of the discrete groups; since, in accordance with the assumption, failure in one course can not mean anything worse than having to repeat the course, and similarly passing can not mean more than that the student is more or less well prepared to begin the next course. This assumption sounds quite absurd when stated thus baldly, yet it is the only assumption that in any way justifies the giving of separate examinations to each year-group. If the assumption were correct, then separate examinations, having the same relations on the scale of difficulty as the different classes have on the scale of achievement, might, presumably, be given with satisfactory results.

It is clear from Chart 16, however, that whatever examinations or administrative procedures are responsible for assigning the students to the year-classes in which they took the Regents examinations in June, 1925, the classes were not formed in accordance with the assumption illustrated in the figure above. The classification is so poor that it is difficult to avoid the conclusion that teaching efficiency must have been seriously impaired, and constructive educational guidance of students rendered practically impossible. Since the sigma-range shows the range of scores of the middle two-thirds of each class, it is easy to see that the upper 15% of French III students are considerably above the French IV average, and that the lower 15% of the French IV students are well below the French III average. And so with the German and Spanish classes.

"School overlapping" distinguished from "Regents overlapping." — These overlappings, however, while they fairly represent the heterogeneity of classes under which both students and teachers labored during the session 1924–1925, are not to be interpreted as the results solely of the Regents examinations. The sigma-ranges in Chart 16 are for *all* the students in each class who took the Regents examinations in June, 1925. We may charge to the Regents examinations only such overlapping as remains *after* these students have been reclassified by the old-type parts of the Regents examinations of June, 1925. We shall present data on such residual overlapping, which is hereafter called "Regents overlapping," after we have analyzed in more detail the "school overlapping" of Chart 16. For this purpose we shall present the distributions of Tables 28–31, inclusive, in the form of percentile graphs.

"School overlapping" displayed by percentile graphs. — The simplest and most effective description of the percentile curve and its uses has been prepared by Otis.[1] "A percentile curve is a smooth line having a hori-

[1] Otis, A. S., *Statistical Method in Educational Measurement*, World Book Company, Yonkers, N. Y., 1925, pp. 53–67. This book cannot be too highly recommended to those who wish to become acquainted with the fundamentals of statistical methods in education. Cf. pp. 19–27, this volume, for confirmatory evidence on overlapping of classes in junior high schools.

zontal length representing 100 per cent of the scores of any group of individuals and so drawn that any point on the curve has a height representing the amount of a given score and a horizontal position on the graph representing the per cent of the scores of the group that is exceeded by the given score. A percentile curve shows at a glance not only the median score of a class but also the range and variability of the scores. It shows at a glance just what per cent of the scores of the class is exceeded by the score of any given individual and just what per cent of the class attains or exceeds any given score. *Two or more curves on the same graph show very vividly the amount of overlapping of the scores of different classes.*"

Charts 17, 18, and 19 are percentile graphs for all the students that took the Regents examinations in French, Spanish, and German in June, 1925, showing percentile curves based on the new-type scores of each class in each language. As a guide to the reader unfamiliar with such graphs we shall "read" in detail Chart 17, showing percentile curves of the new-type scores of second-, third-, and fourth-year French classes.

How to read a percentile graph.[1] — The lowest percentile curve in Chart 17 is for French II, and the upper for French IV students. The first thing to notice, and keep in mind, is that the point at which a curve crosses the fifty-percentile line, i.e., the heavy vertical line headed "50," represents the median score of the class; the value of this median score can be ascertained from the first column of figures at the left, headed "New-type scores." The French II curve crosses the fifty-percentile line just above the horizontal line corresponding to a score of 150; since the distance between each pair of horizontal lines represents 10 score-points, and the French II curve crosses at a point about two-tenths of the distance between the 150- and 160-score horizontal lines, we estimate that the median of the French II class is about 152 new-type score-points. Similarly we may estimate the French III median to be about 184, and the French IV median to be about 208 score-points. (Since the distributions are slightly skewed, the medians differ slightly from the mean scores indicated in Chart 16.)

One-third of students in French classes misplaced by one year or more. — Applying a straight-edge, which is kept constantly parallel to the horizontal lines in the chart, to the point at which the French II curve crosses the fiftieth-percentile line, we may ascertain the per cents of French III and French IV students who receive scores below the French II average by noting the points at which the straight-edge cuts the French III and IV curves, and reading the per cents vertically above these points. The straight-edge when so applied cuts the French III curve just to the right of the 10th-percentile (vertical) line, and cuts the French IV curve just

[1] See above, pp. 20 to 27.

New Type Scores	French II		French III		French IV	
250			6741	100.0	489	100.0
240	13486	100.0	6734	99.9	483	98.8
230	13480	99.9	6699	99.3	447	91.4
220	13457	99.8	6560	97.3	398	81.4
210	13378	99.2	6259	92.8	334	68.3
200	13181	97.8	5726	84.9	262	53.6
190	12757	94.6	4900	72.7	187	38.2
180	12022	89.2	3900	57.9	133	27.2
170	11001	81.6	2831	42.0	96	19.6
160	9710	72.0	1910	28.3	58	11.9
150	8097	60.0	1158	17.2	27	5.5
140	6406	47.5	708	10.5	13	2.7
130	4829	35.8	413	6.1	6	1.2
120	3367	25.0	230	3.4	4	.8
110	2177	16.1	130	1.9	3	.6
100	1389	10.3	79	1.2	2	.4
90	834	6.2	50	.7	1	.2
80	482	3.6	30	.4		
70	252	1.9	22	.3		
60	141	1.0	17	.3		
50	65	.5	10	.1		
40	34	.3	4	.06		
30	15	.1	1	.01		
20-29	5	.04	1	.01		
10-19	1	.008				

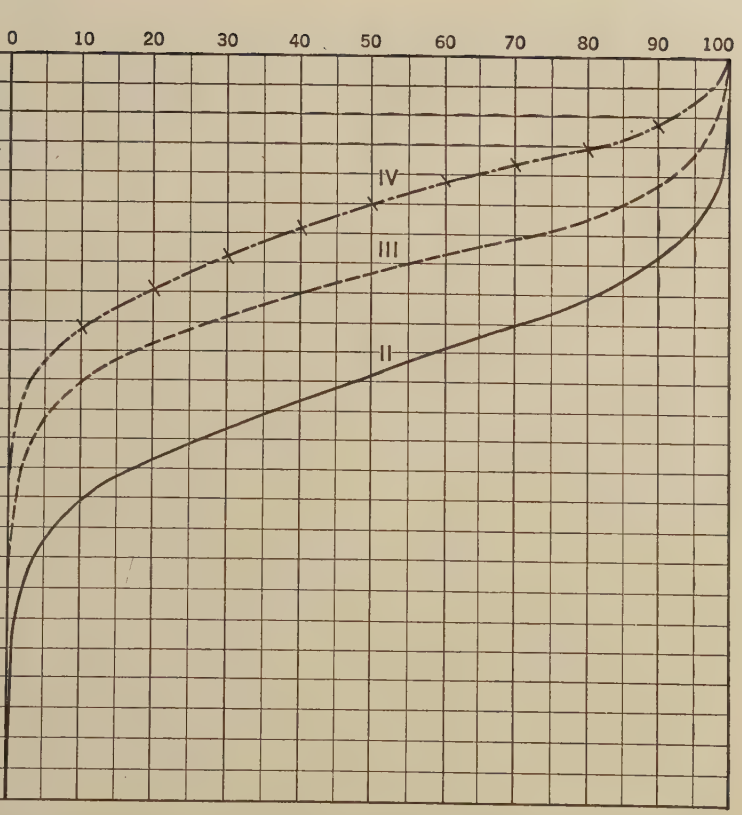


CHART 17. — Percentile curves of new-type scores of all 2nd-, 3rd-, and 4th-year French students who took the Regents examinations in June, 1925.

to the left of the 5th-percentile line; we thus perceive that about 12% of the French III and about 4% of the French IV students are below the French II average. It appears, therefore, that of the 489 students who had studied French in New York State high schools five periods per week for four years, about twenty are so far from being fourth-year students in actual achievement that they would be scarcely respectable, academically, as members of the second-year class! How these students were ever admitted to the fourth-year class is a mystery which we shall endeavor to explain in a later section of this report.

Applying the straight-edge to the point at which the French III curve crosses the fiftieth-percentile (vertical) line, and keeping it constantly parallel to the horizontal lines of the chart, we observe that it cuts the French II curve on the eighty-fifth-percentile line; subtracting 85 from 100, we learn that about 15% of the French II students are above the average of the French III class. Holding the straight-edge in this position, we observe (at the left) that it cuts the French IV curve just to the right of the twentieth-percentile line; whence we estimate that about 22% of the French IV students are below the French III average in actual achievement in French.

Keeping the straight-edge constantly parallel to the horizontal lines of the chart and sliding it up to the point at which the French IV curve crosses the fiftieth-percentile line, we note that the straight-edge cuts the French III curve just to the right of the eightieth-percentile line, and cuts the French II curve just to the right of the ninety-fifth-percentile line. Estimating the exact per cents, and subtracting from 100, it appears that about 17% of the French III and about 3% of the French II students are above the fourth-year average. We have no means of directly measuring the misplacement of students at the lower end of the French II and at the upper end of the French IV curves; but since about 22% of the French IV students are below, and about 15% of the French II students are above the French III median, and since only about 70% of the French III students are between the French II and French IV medians, it seems fair to say that at least 30% of all French students are misplaced in one direction or the other by one year or more, and about 3 or 4% by two years or more.

Sixty per cent misplaced by as much as one semester. — What proportion of the students are misplaced by as much as one semester? We may answer this question by applying the straight-edge to the point midway between the points at which the French II and III curves cross the fiftieth-percentile line, and to the point midway between the French III and IV median points. The median for French II is 152 and for French III 184; the point midway between these two is 168. Applying the straight-edge to this point while keeping it parallel to the horizontal lines, we learn

New Type Scores	Spanish II		Spanish III		Spanish IV	
260			2132	100.0	226	100.0
250	5340	100.0	2130	99.9	224	99.1
240	5339	99.9	2120	99.4	212	93.8
230	5324	99.7	2083	97.7	178	78.8
220	5286	99.0	1977	92.7	135	59.7
210	5193	97.2	1847	86.6	97	42.9
200	4998	93.6	1624	76.2	60	26.5
190	4712	88.3	1315	61.7	34	15.0
180	4259	79.7	951	44.6	17	7.5
170	3675	68.8	622	29.2	7	3.1
160	3082	57.8	374	17.5	2	.9
150	2394	44.9	197	9.2		
140	1745	32.7	86	4.0		
130	1136	21.3	41	1.9		
120	735	13.8	23	1.1		
110	416	7.8	15	.7		
100	235	4.4	6	.3		
90	113	2.1	2	.1		
80	58	1.1	1	.05		
70	33	.6	1	.05		
60	15	.3				
50-59	6	.1				
40-49	1	.02				

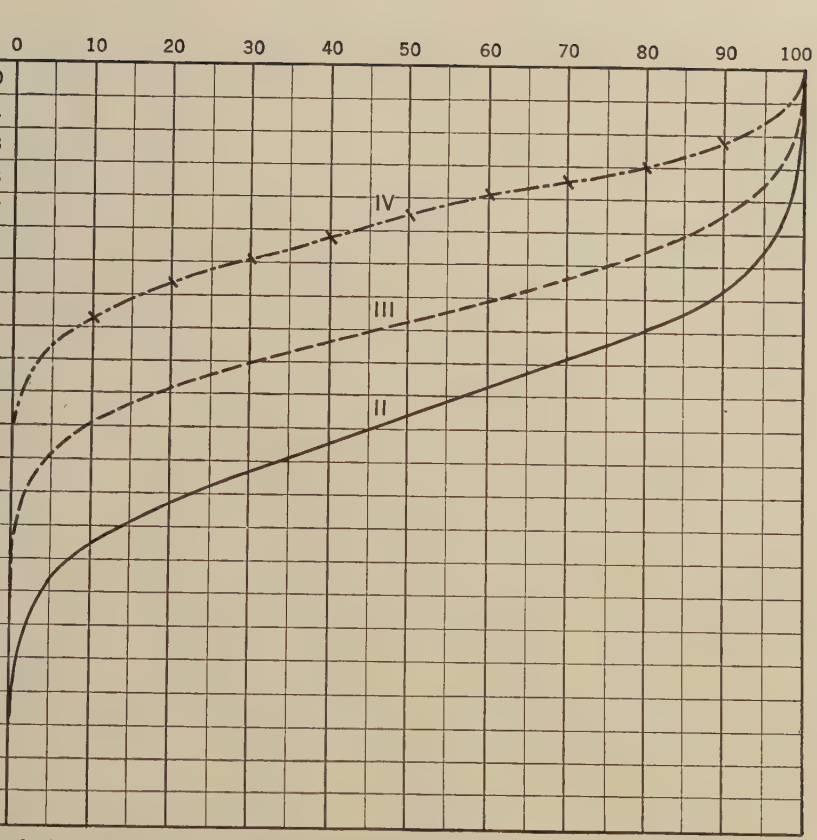


CHART 18. — Percentile curves of new-type scores of all 2nd-, 3rd-, and 4th-year Spanish students who took the Regents examinations in June, 1925.

New Type Scores	German II		German III		German IV	
270			667	100.0		
260	1482	100.0	666	99.8	127	100.0
250	1480	99.9	659	98.8	126	99.2
240	1474	99.5	624	93.5	118	92.9
230	1451	97.9	572	85.8	103	81.1
220	1404	94.8	505	75.7	79	62.2
210	1329	89.7	443	66.4	65	51.2
200	1236	83.4	371	55.6	37	29.1
190	1095	73.9	297	44.5	27	21.3
180	944	63.7	218	32.7	16	12.6
170	790	53.3	163	24.4	10	7.9
160	664	44.8	109	16.4	6	4.7
150	527	35.6	76	11.4	5	3.9
140	400	27.0	39	5.8	2	1.6
130	309	20.9	26	3.9		
120	240	16.2	19	2.8		
110	163	11.0	13	1.9		
100	116	7.8	6	.9		
90	74	5.0				
80	38	2.6				
70	20	1.3				
60-69	10	.7				
50-59	4	.3				

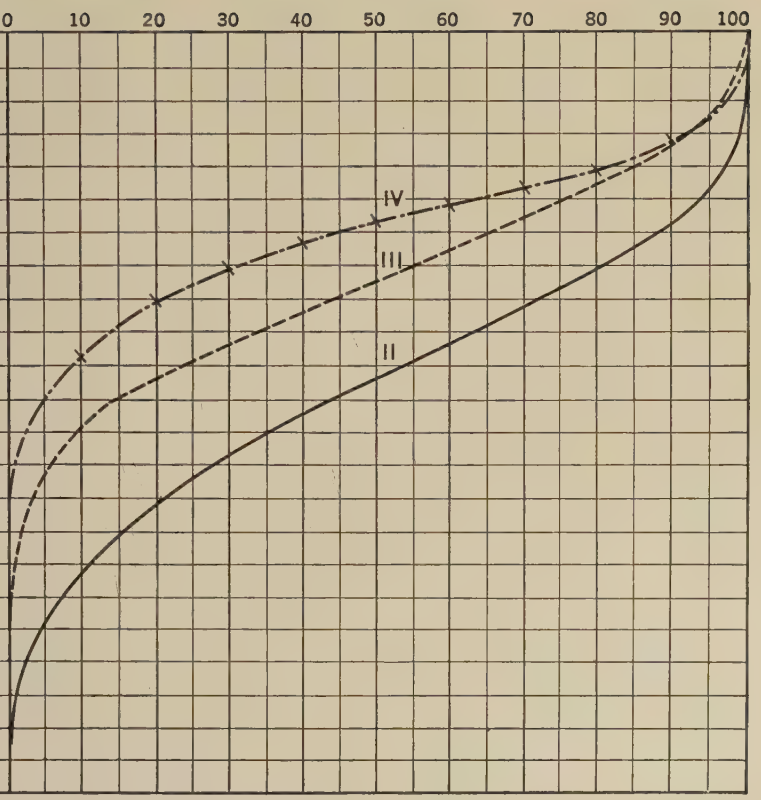


CHART 19. — Percentile curves of new-type scores of all 2nd-, 3rd-, and 4th-year German students who took the Regents examinations in June, 1925.

by reading the per cents vertically above the points at which it cuts the French IV, III and II curves, that about 11% of the French IV students are nearer the French II median than to a higher class median; that about 25% of the French III students are nearer to the French II median than to a higher median; and that about 32% of French II students are nearer to a higher median than to their own class median.

Sliding the straight-edge up to the point midway between the median points of French III and IV students, and reading from left to right above the points at which it cuts the three curves, we learn that about 33% of French IV students are nearer to a lower class median than to their own, that about 35% of French III, and about 9% of French II students are nearer to the French IV median than to a lower class median. Of the French III students, therefore, about 65% are nearer to a lower or higher class median than to their own; less than 40% are closer to the French III median than to some other class median! From these facts it seems safe to conclude that more than 60% of all French students in New York State secondary schools are misplaced by one semester or more. This condition is very near to chaos, so far as classification for instructional purposes and so far as educational guidance are concerned. It is an eloquent portrayal of the crucial need for a centrally administered system of examinations capable of affording accurate measures of defined achievement expressed in terms of comparable units and reckoned from stable and uniform standards. The overlapping here displayed is what we have called "school overlapping"; obviously, any examination system which depends on local ratings from schools must accept such overlapping of classes and misplacement of students *to the full extent of such dependence.*

Charts 18 and 19, showing percentile graphs for Spanish and German classes, are in every respect similar to Chart 17, and may be read in the same way. They are, therefore, presented without comment.

Regents overlapping. — In order to measure the overlapping caused, or allowed, by the Regents examinations we must consider only those students who, according to old-type Regents examinations of June, 1925, belong in each of the three classes; that is, those students in each class who are passed by the Regents marks or by marks which the State Department readers have accepted. Charts 20, 21, and 22 show percentile curves for each of the three classes in each modern language. On account of space limitations we shall describe only the percentile graph for the French classes, leaving it to the reader to see how closely the indications of Charts 21 and 22 parallel those of Chart 20.

Percentile graphs for students passed by Regents review ratings. — The points at which the three curves cross the fiftieth-percentile line in Chart 20 indicate that the median scores of the passed students of the three French classes are about 160, 189, and 213 in order. Applying a straight-edge

New Type Score	French II		French III		French IV	
250			5717	100.0	390	100.0
240	10291	100.0	5710	99.9	384	98.5
230	10285	99.9	5675	99.3	348	89.3
220	10262	99.8	5538	96.9	300	77.0
210	10184	99.0	5238	91.6	240	61.5
200	9987	97.1	4708	82.4	179	45.9
190	9563	93.0	3893	68.1	110	28.2
180	8835	85.9	2923	51.2	71	18.2
170	7832	76.1	1935	33.9	48	12.3
160	6583	64.0	1146	20.1	26	6.7
150	5069	49.3	556	9.7	8	2.1
140	3577	34.8	263	4.6	3	.8
130	2302	22.4	115	2.0	2	.5
120	1285	12.5	40	.7	2	.5
110	616	6.0	13	.2	2	.5
100	290	2.8	4	.07	2	.5
90	112	1.1			1	.3
80	40	.4				
70	9	.09				
60	3	.03				
50						
40						
30						
20-29						
10-19						

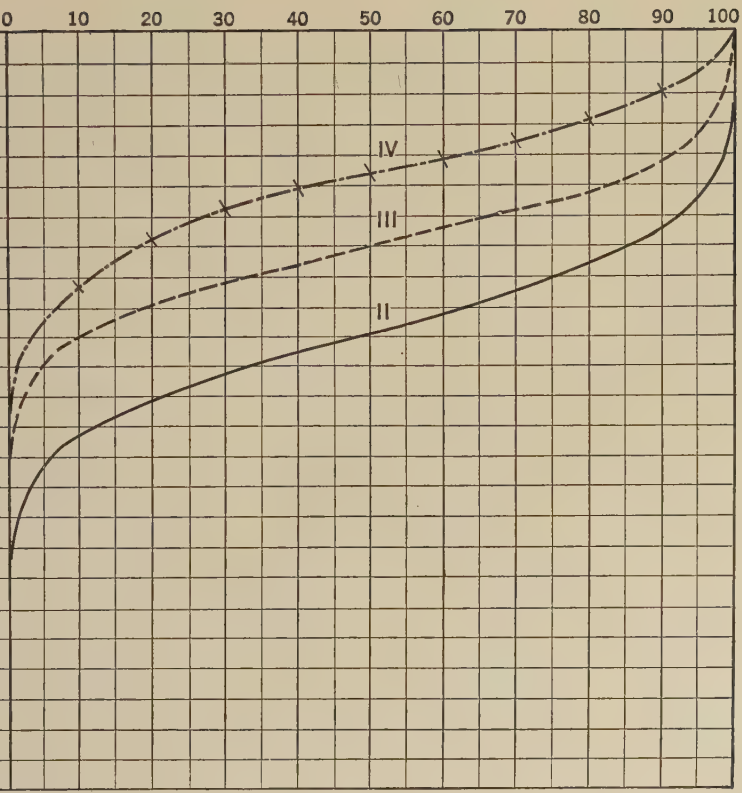


CHART 20. — Percentile curves of new-type scores of all 2nd-, 3rd-, and 4th-year students who were passed by the old-type parts of the Regents French examinations in June, 1925.

New Type Scores	Spanish II		Spanish III		Spanish IV	
260			1719	100.0	210	100.0
250	4282	100.0	1717	99.9	208	99.1
240	4281	99.9+	1707	99.4	196	93.4
230	4266	99.6	1671	97.3	162	77.1
220	4228	98.7	1566	91.2	120	57.2
210	4137	96.6	1436	83.6	84	40.0
200	3943	92.1	1218	70.9	50	23.8
190	3659	85.4	919	53.5	26	12.4
180	3214	75.1	583	33.9	13	6.2
170	2638	61.6	398	23.2	5	2.4
160	2075	48.5	203	11.8	1	.5
150	1453	33.9	87	5.1		
140	912	21.3	25	1.5		
130	461	10.8	11	.6		
120	220	5.1	3	.2		
110	83	1.9	2	.1		
100	27	.6				
90	8	.2				
80	4	.1				
70	2	.05				
60	1	.02				
50-59	1	.02				
40-49	1	.02				

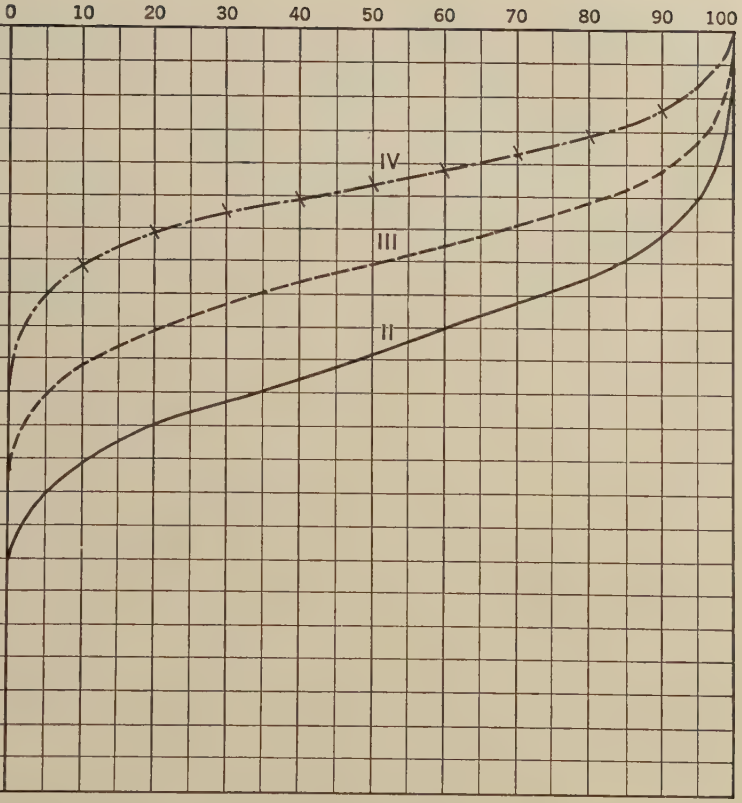


CHART 21. — Percentile curves of new-type scores of all 2nd-, 3rd-, and 4th-year students who were passed by the old-type parts of the Regents Spanish examinations in June, 1925.

New Type Scores	German II		German III		German IV	
270			581	100.0		
260	1336	100.0	580	99.8	110	100.0
250	1334	99.8	573	98.6	109	99.2
240	1328	99.4	538	92.6	101	91.2
230	1305	97.7	486	83.7	86	78.2
220	1258	94.2	419	72.1	62	56.4
210	1183	88.5	357	61.4	49	44.5
200	1090	81.6	287	49.4	22	20.0
190	949	70.3	216	37.2	14	12.7
180	800	59.9	149	29.6	4	3.6
170	649	48.6	98	16.9	2	1.8
160	526	39.4	53	9.1	1	.9
150	398	29.8	34	5.9	1	.9
140	284	21.3	13	2.2		
130	203	15.2	10	1.7		
120	146	10.9	9	1.5		
110	83	6.2	6	1.0		
100	52	3.9	2	.3		
90	30	2.2				
80	11	.8				
70	3	.2				
60-69	1	.07				
50-59	1	.07				

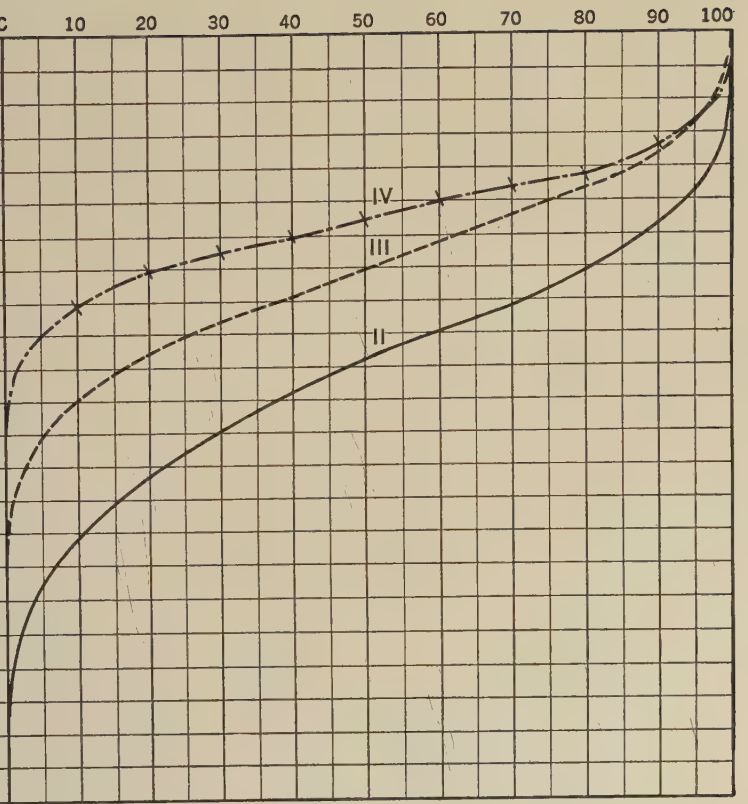


CHART 22. — Percentile curves of new-type scores of all 2nd-, 3rd-, and 4th-year students who were passed by the old-type parts of the Regents German examinations in June, 1925.

to the point at which the French II curve crosses the fiftieth-percentile line, keeping it parallel to the horizontal lines of the chart, and noting the points at which it cuts the other two curves at the left, we learn that (in terms of new-type scores) about 10% of the students which the Regents examinations classify as having passed French III are below the average of those classified by the old-type Regents examinations as having passed French II, and that about 2% of the students passed in French IV, and thus declared by the Regents examinations to be prepared to study French literature in college, are below the average of students passed in French II! It is thus manifest that while the Regents examinations do somewhat reduce the school overlapping, they fall far short of classifying and promoting students according to actual achievements. The Regents examinations promote to college classes in French literature some students who are unable to recognize the meanings of French words, to read simple French sentences, or to complete very simple French phrases, as well as these things are done by the average students who, according to the Regents examinations, have passed French II. One in ten of the students promoted by the old-type Regents examinations to French IV is below the average of those promoted only to French III. This failure of the Regents examinations to do the very thing for which they were primarily instituted, and which is so essential to the educational progress of the state, is inherent in the form of examinations used, and is eloquent of the need for fundamental changes in the Regents examination system.

Applying the straight-edge to the point at which the French III curve crosses the fiftieth-percentile line and reading the per cents vertically above the points at which it cuts the French IV curve at the left, and the French II curve at the right we find that about 18% of students passed in French IV are below, and about 16% of the students passed in French II are above the median of students passed in French III. Sliding the straight-edge up to the French IV median point and reading vertically above the points at which it cuts the French II and III curves, it appears that about 14% of French III passed students and about 2% of French II passed students are above the average of students passed by the old-type part of the June, 1925, Regents examination in French IV.

Applying the straight-edge to the point on the fiftieth-percentile line midway between the French II and French III medians, keeping it parallel to the horizontal lines, and reading above the points where it cuts the French III and French IV curves, we learn that about 8% of the passed French IV students are nearer the French II median than to a higher median, that is, are misplaced by at least one year or more; and that about one-fourth of passed French III students are nearer to the French II median than to their own class median. Holding the straight-edge in this position and reading at the right, we learn that about 33% of passed

French II students are nearer to the average of passed French III or French IV students than to their own class average.

Sliding the straight-edge up to the point midway between the French III and IV median points and reading at the left we learn that about 30% of French IV students passed by the old-type Regents examination are nearer to a lower class median than to their own; and, reading at the right, we learn that about 32% of passed French III students and about 8% of passed French II students are nearer to the French IV median than to any other class median.

Regents overlapping nearly equal to school overlapping. — Summarizing, it appears that by the old-type Regents examinations in French about five or six per cent of all French students are misplaced by two years or more, about 33% by one year or more, and more than 60% by one semester or more. Less than 40% of the students as classified by the Regents old-type examinations in French are nearer to their own class median than to some other class median. On the whole, this is almost as bad as the classification based on local school ratings of the old-type Regents examinations.

That such failure to classify reasonably well is thoroughly characteristic of the Regents examination system as it operates with old-type examinations is indicated by Charts 21 and 22, which show similar overlappings for Spanish and German classes. These charts lend almost incontrovertible support to the analysis of old-type examinations offered above on pages 133 to 136.

Percentile graphs for students failed by Regents review ratings. — One would naturally expect that most of the students failed by the old-type Regents examinations would be at the lower end of the distribution of new-type scores in each class. The amount of "Regents overlapping" just described shows that, although a sufficient number was failed in each class to reduce materially the overlapping, the old-type Regents examinations actually did not distribute the failures in such a way as to improve very much on the poor classification effected by the local school ratings. Charts 23, 24, and 25 show strikingly the indiscriminate character of the old-type failures.

A full appreciation of Chart 23 requires that it be compared in detail with Charts 17 and 20; but in itself it presents a striking picture. It is clear from Chart 23 that many students who are failed in French are above the average of the class in which they were failed and in a few cases are even above the average of the class above that in which they were failed. For example, if the reader will recall that the median of all French IV students was 208, Chart 23 shows that a little over 1% of the students failed in French III are above the average score of all French IV students. Even more striking is the fact that a few students were failed in

New Type Score	French II		French III		French IV	
250						
240						
230			1024	100.0	99	100.0
220	3095	100.0	1022	99.8	98	99.0
210	3094	99.9	1021	99.7	94	95.0
200	3094	99.9	1018	99.5	83	84.0
190	3094	99.9	1007	98.4	77	77.8
180	3087	99.8	977	95.4	62	62.6
170	3069	99.2	896	87.5	48	48.5
160	3027	97.8	764	74.6	32	32.3
150	2928	94.6	602	58.8	19	19.2
140	2729	88.2	445	43.5	10	10.1
130	2427	78.4	298	29.1	4	4.0
120	1982	64.0	190	18.6	2	2.0
110	1461	47.2	117	11.4	1	1.0
100	999	32.3	75	7.3		
90	622	20.1	50	4.9		
80	342	11.0	30	2.9		
70	243	7.9	22	2.1		
60	138	4.5	17	1.7		
50	65	2.1	10	1.0		
40	34	1.1	4	.4		
30	15	.5	1	.1		
20-29	5	.2	1	.1		
10-19	1	.03				

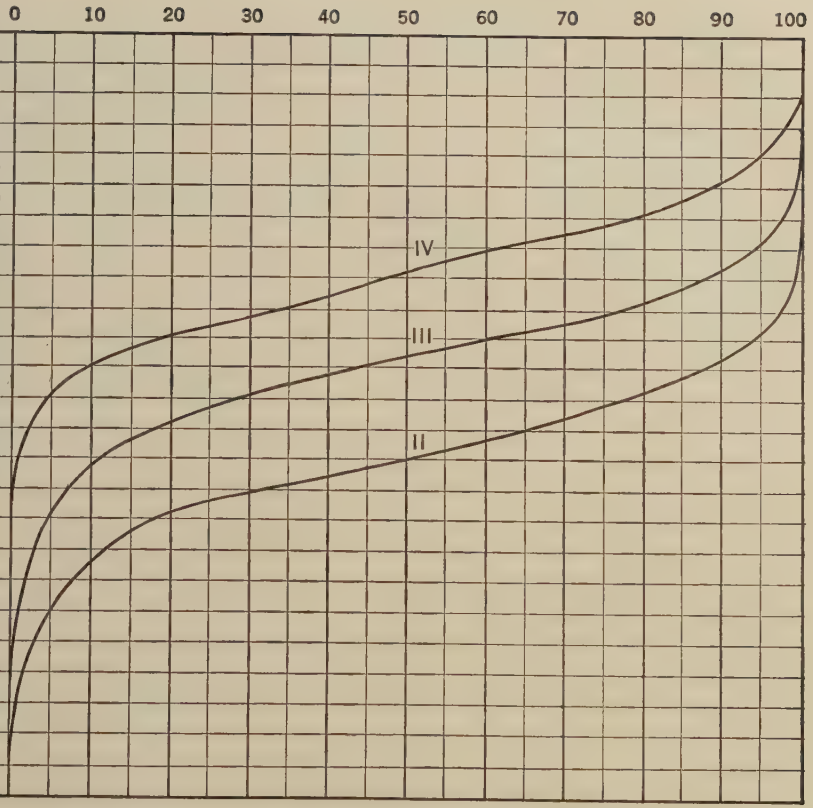


CHART 23. — Percentile curves of new-type scores of all 2nd-, 3rd-, and 4th-year French students failed by the old-type parts of the Regents French examinations in June, 1925.

New Type Scores	Spanish II		Spanish III		Spanish IV	
260						
250						
240			313	100.0		
230			312	99.7	16	100.0
220	1058	100.0	311	99.4	15	93.8
210	1056	99.8	311	99.4	13	81.0
200	1055	99.8	306	97.8	10	62.5
190	1053	99.6	296	94.6	8	50.0
180	1045	98.8	268	85.6	4	25.0
170	1037	98.1	224	71.6	2	12.5
160	1007	95.2	171	54.7	1	6.3
150	941	89.0	110	35.1		
140	833	78.8	61	19.5		
130	675	63.8	30	9.6		
120	515	48.7	20	6.4		
110	333	31.5	13	4.2		
100	208	19.7	6	1.9		
90	105	9.9	2	.6		
80	54	5.1	1	.3		
70	31	2.9	1	.3		
60	14	1.3				
50-59	5	.5				
40-49						

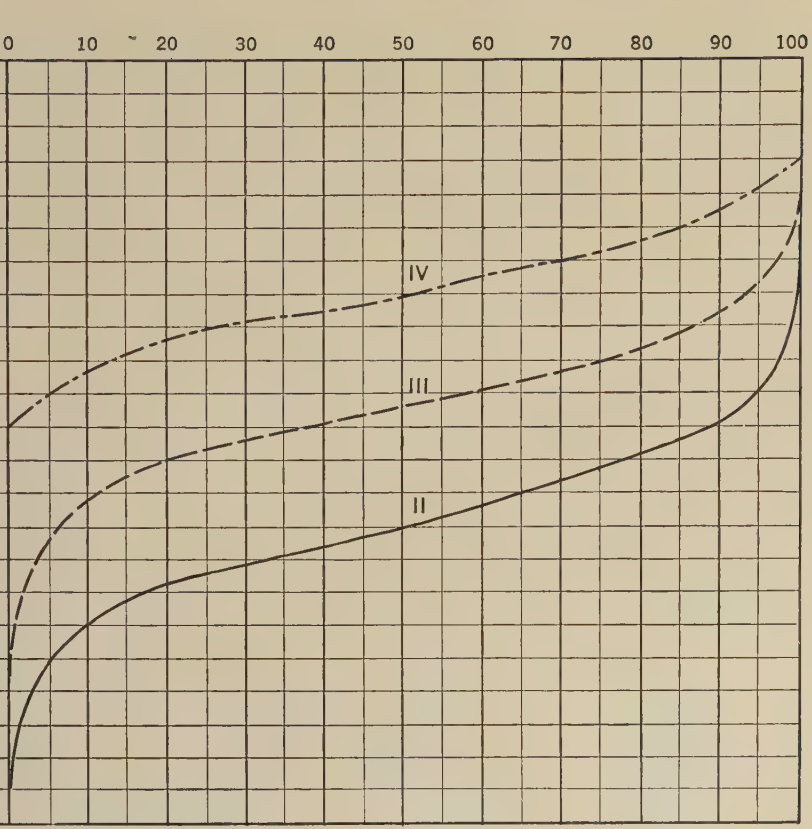


CHART 24. — Percentile curves of new-type scores of all 2nd-, 3rd-, and 4th-year Spanish students failed by the old-type parts of the Regents Spanish examinations in June, 1925.

New Type Score	German II		German III		German IV	
270						
260						
250						
240						
230						
220					17	100.0
210			86	100.0	16	94.2
200			84	97.7	15	88.3
190	146	100.0	81	94.2	13	76.5
180	144	98.7	69	80.3	12	70.6
170	141	96.6	65	75.6	8	47.1
160	138	94.6	56	65.1	5	29.4
150	129	88.4	42	48.9	4	23.5
140	116	79.5	26	30.2	2	11.8
130	106	72.6	16	18.6		
120	94	64.4	10	11.6		
110	80	54.8	7	8.1		
100	64	43.9	4	4.7		
90	44	30.1				
80	27	18.5				
70	17	11.6				
60-69	9	6.2				
50-59	3	2.1				

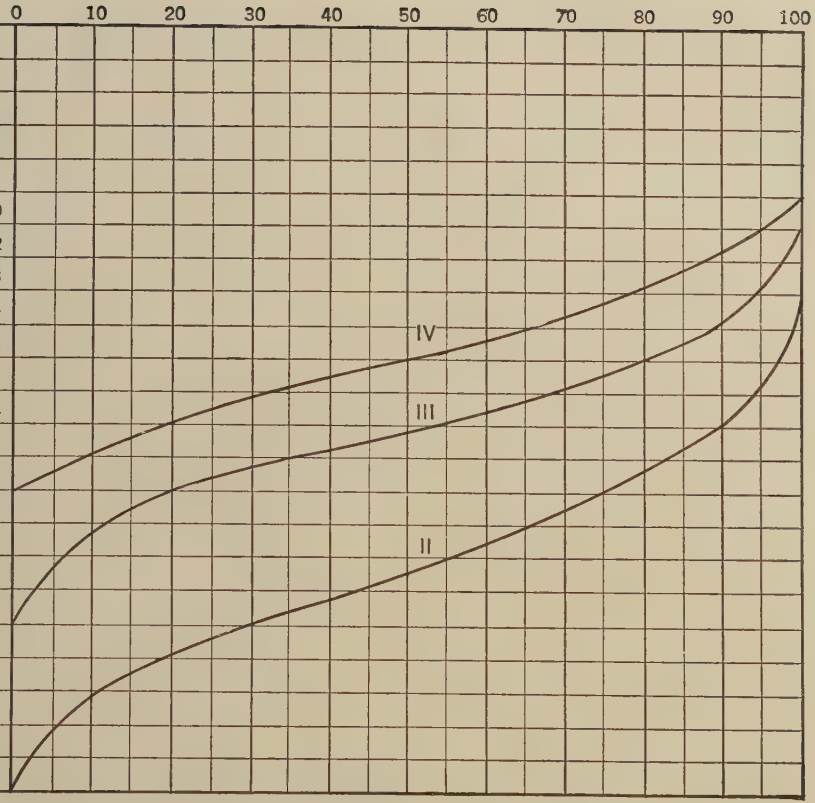


CHART 25. — Percentile curves of new-type scores of all 2nd-, 3rd-, and 4th-year German students failed by the old-type parts of the Regents German examinations in June, 1925.

French II who were above the French IV average. The educational significance of such errors in grading can hardly be overestimated. Charts 24 and 25 are similar to Chart 23 and indicate that the differences between the three languages are so slight as to be negligible. The educational welfare of the students in all three languages demands that the Regents examinations be materially strengthened. Table 36 is presented as a compact numerical summary of the main aspects of the overlapping set forth in Charts 17–25, inclusive.

TABLE 36

OVERLAPPING OF MODERN LANGUAGE CLASSES, SHOWING THE PER CENT OF EACH CLASS THAT EXCEEDS THE NEXT HIGHER CLASS AVERAGE, AND THE PER CENT THAT FALLS SHORT OF THE NEXT LOWER CLASS AVERAGE ACCORDING TO NEW-TYPE SCORES. THE PER CENTS IN ODD-NUMBERED COLUMNS APPLY ONLY TO STUDENTS PASSED BY THE DEPARTMENT; THE OTHER FIGURES RELATE TO ALL STUDENTS THAT TOOK THE NEW-TYPE TESTS IN EACH CLASS IN JUNE, 1925. THE CLASS MEANS USED IN EVEN-NUMBERED COLUMNS ARE OF THE TOTAL NUMBERS OF STUDENTS IN EACH CLASS, AND IN ODD-NUMBERED COLUMNS OF STUDENTS PASSED BY THE DEPARTMENT REVIEWERS.

	PER CENTS OF STUDENTS IN CLASSES INDICATED IN COLUMN (1) WHO ARE							
	Above 3rd-Year Mean		Above 4th-Year Mean		Below 2nd-Year Mean		Below 3rd-Year Mean	
	Total	Passed	Total	Passed	Total	Passed	Total	Passed
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
French II	15.0	15.0	4.0	2.0				
III			22.0	14.8	10.0	9.7		
IV					3.0	2.0	22.0	17.5
Spanish II	18.0	16.4	2.0					
III			10.0	14.0	12.0	13.0		
IV						1.0	10.0	11.0
German II	23.0	18.4	6.0	7.0				
III			45.0	30.0	18.0	18.0		
IV					6.0	2.0	24.0	23.6

Percentile graph for new-type Physics test scores of total, passed and failed groups. — Chart 26 shows percentile curves for all the students who took the Regents examinations in physics in June, 1925; for all students who were passed on the old-type part of the Regents examination in physics, and for all students who were failed by the old-type part of the Regents examination. The inaccuracy of the grades based on the old-type part of the Regents examination is clearly indicated. Many students who are far above the average are failed by the old-type, while many students who are obviously failures are given credit by the old-type examination.

Score	Total Subtotal	Total Percent	Passed Subtotal	Passed Percent	Failed Subtotal	Failed Percent
15	14081	100.0	11717	100.0		
20	14078	99.9+	11714	99.9+		
25	14071	99.9	11707	99.9		
30	14049	99.7	11685	99.8		
35	14006	99.5	11642	99.4		
40	13953	99.1	11589	98.9	2364	100.0
45	13843	98.4	11480	98.0	2363	99.9+
50	13701	97.4	11339	96.8	2362	99.9
55	13460	95.6	11100	94.8	2360	99.8
60	13182	93.7	10824	92.8	2358	99.7
65	12751	90.6	10398	88.7	2353	99.6
70	12286	87.3	9942	84.9	2344	99.2
75	11632	82.6	9301	79.4	2331	98.7
80	10919	77.6	8612	73.5	2307	97.6
85	9975	70.8	7717	65.8	2258	95.5
90	9007	64.0	6819	58.2	2188	93.8
95	7871	55.9	5776	49.3	2095	88.6
100	6649	47.2	4672	39.9	1977	83.6
105	5409	38.4	3636	31.0	1773	75.0
110	4306	30.6	2749	23.4	1557	65.9
115	3221	22.9	1923	16.4	1298	54.9
120	2385	16.9	1346	11.5	1039	44.0
125	1567	11.1	827	7.1	740	31.3
130	1004	7.1	479	4.1	525	22.2
135	544	3.9	227	1.9	317	13.4
140	270	1.9	105	.9	165	7.0

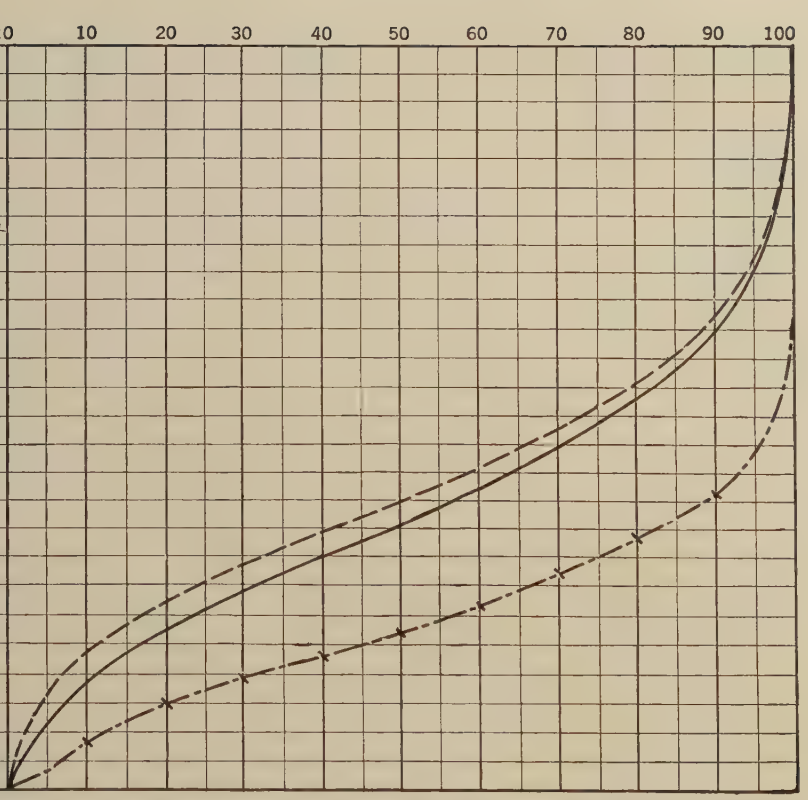


CHART 26. — Percentile curves of new-type physics test scores of 14,081 students who took the Regents physics examination in June, 1925, of 11,717 who passed, and of 2364 who failed the old-type part of the examination. The upper curve represents passed students, and the lower failed students.

Overlapping of classes displayed by frequency polygons. — Thus far we have attempted to show the character and extent of overlapping of classes due to the inaccuracies of the old-type parts of the Regents examinations by means of percentile curves. While the percentile curve is peculiarly adapted to display overlapping, it is by no means the only device by which extensive overlapping may be exhibited. This is especially true for readers who are not used to reading percentile curves. For the sake of all readers of this report, but especially for those who may be unfamiliar with the percentile graph, Charts 27 to 29 inclusive are presented.

The frequency polygons of Chart 27 display in a most striking manner the overlapping of classes before and after the application of the old-type parts of the Regents examination in French. The line of figures at the bottom of the chart reading from left to right refers to scores on the new-type test in French. The figures reading up at the left represent the numbers of students. The curves at the bottom of Chart 27 are for all students in the second-, third-, and fourth-year classes, respectively, who took the Regents examinations in June, 1925. The solid line curve shows that the scores of the second-year students range from 10 to 240, those of third-year students from 20 to 250, and those of fourth-year students from 90 to 250. The vertical lines show the means or averages of the three classes.

The curves at the top refer only to students in these three classes who passed the old-type part of the Regents examinations in French in June, 1925. The very slight effect which the old-type Regents examination had on the forms of the total distributions and on the average scores of the three classes is self-evident in Chart 27. The spread of the scores in the three classes is scarcely reduced at all and the overlapping of classes, as has already been set forth in connection with Charts 20, 21, and 22, is only very slightly reduced.

Charts 28 and 29 are very similar to Chart 27 and should be read in the same manner.

In order to make the facts of overlapping of classes as clear as possible and to show the evidence on the effectiveness or ineffectiveness of the old-type Regents examinations in as many lights as possible, we present the facts of the preceding charts in another series of frequency polygons in Charts 30 to 33.

The advantage of Chart 30 is that corresponding groups in the three classes may be directly compared. The three frequency polygons at the top of Chart 30 refer to total, passed and failed groups of French II students. The curves at the bottom of Chart 30 refer to total, passed and failed groups of French IV students. All nine frequency polygons of Chart 30 are set up over the same scale of scores, as indicated under the

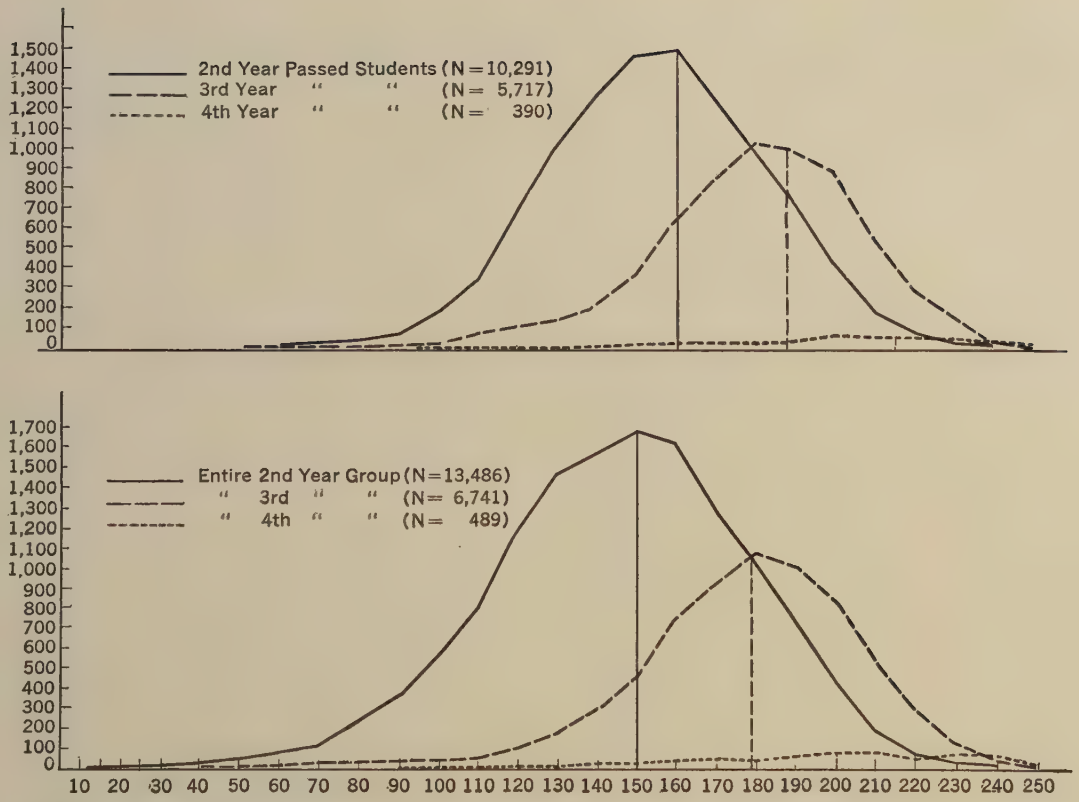


CHART 27. — Graphs of distributions of new-type French examination scores, of all students who took the Regents examinations in June, 1925, in French II, III, and IV, below; and above, of all students in the same classes who passed the old-type part of the Regents examinations. The scores for both sets of curves are indicated at the bottom, from left to right; the numbers of students are indicated in two scales at the left. The vertical lines represent the medians of the distributions.

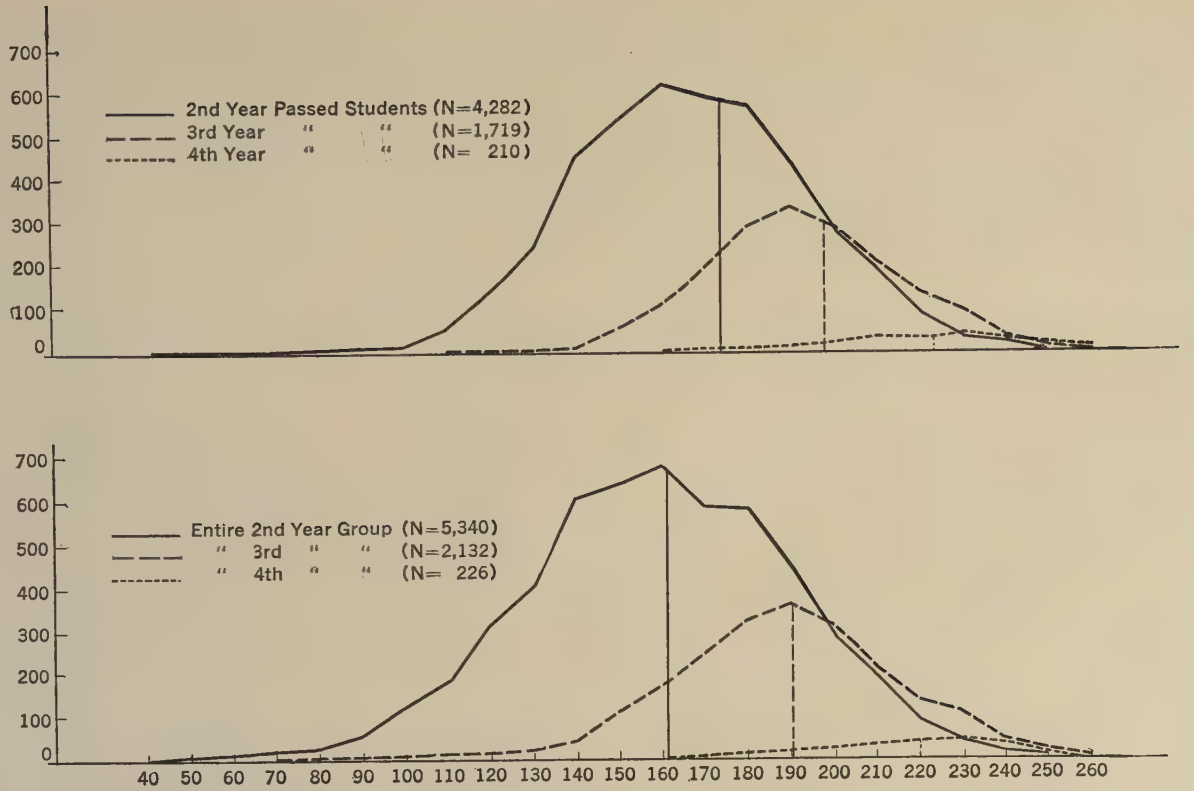


CHART 28. — Graphs of distributions of new-type Spanish examination scores, of all students who took the Regents examinations in June, 1925, in Spanish II, III, and IV, below; and above, of all students in the same classes who passed the old-type parts of the Regents examinations. The scores for both sets of curves are indicated at the bottom, from left to right; the numbers of students are indicated in two scales at the left. The vertical lines represent the medians of the distributions.

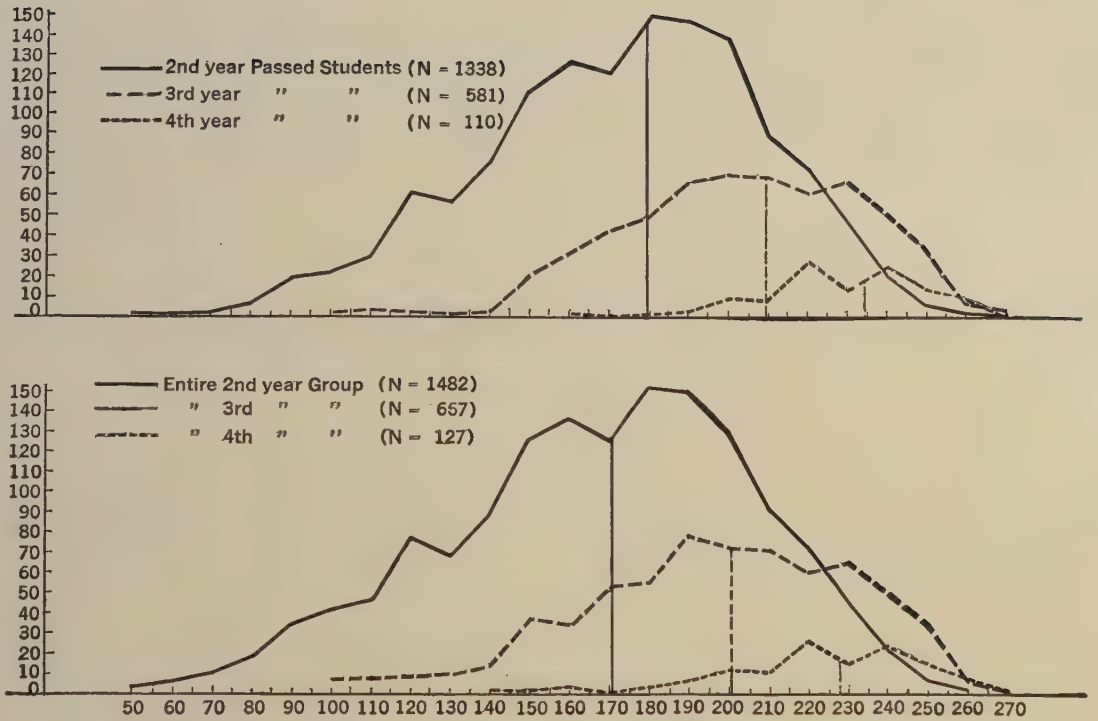


CHART 29. — Graphs of distributions of new-type German examination scores, of all students who took the Regents examinations in June, 1925, in German II, III, and IV, below; and above, of all students in the same classes who passed the old-type parts of the Regents examinations. The scores for both sets of curves are indicated at the bottom, from left to right; the numbers of students are indicated in two scales at the left. The vertical lines represent the medians of the distributions.

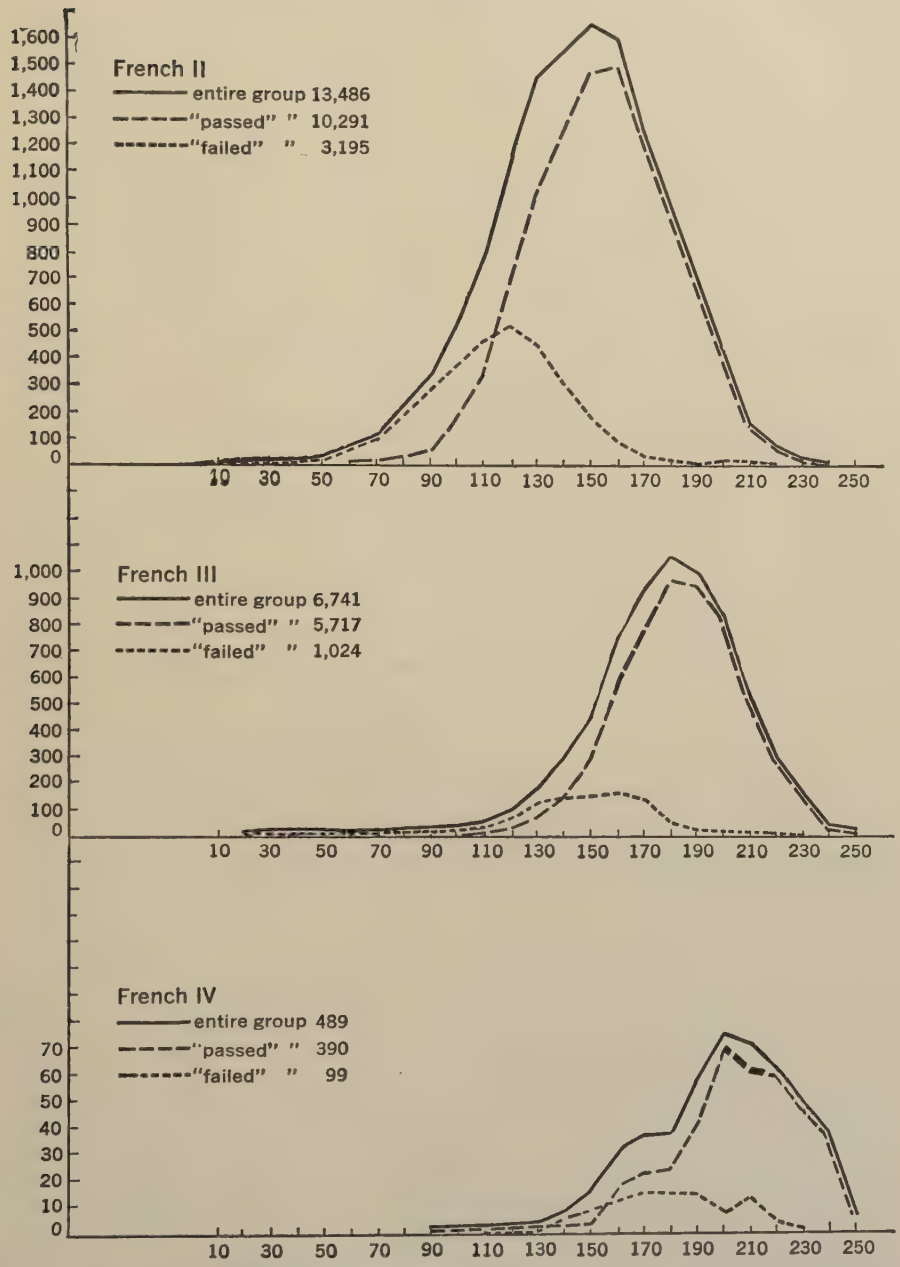


CHART 30. — Graphs of distributions of new-type scores of total population, passed and failed groups of each French class. The passed and failed groups consist of students passed or failed by the old-type parts of the Regents examinations. Scores are represented on the abscissa.

base lines of the polygons. By placing a straight-edge vertically on the chart and sliding it from left to right the reader may make the same comparisons as were made by means of the percentile curves described above. For example, it is obvious from Chart 30 that at least a few students who were failed in second-year French received higher scores on the objective test than half of the students who were passed in fourth-year French received. It is also clear from Chart 30 that some students who were passed in fourth-year French received scores lower than the average score received by students who were failed in second-year French. This is evidence of the signal failure of the Regents examination system which cannot by any reasonable means be controverted. Even though the reader may still be doubtful of the validity of the new-type examinations, the differences in standards set forth by Chart 30 are so great that they cannot be explained away by adverting to the possible lack of validity in the new-type scores. In other words, the force of the evidence here presented does not depend upon an acceptance of the new-type scores as completely valid measures of modern language achievement. This fact is mentioned at this time merely for the sake of clearness, because it is scarcely conceivable that any one who has examined all the data presented above can still entertain any reasonable doubts about the validity of the new-type examinations. If any such doubts do linger in the minds of some readers, it is confidently believed that they will be dispelled before the end of this report is reached.

Charts 31 and 32 are similar to Chart 30 and should be read in the same manner. Chart 33 shows frequency polygons for total, passed and failed groups of students who took the Regents examination in physics in June, 1925. The flattest curve of the three is a graph of the distribution of new-type scores of students failed by the old-type part of the Regents examination. It shows that some students who were in the highest 10% of the whole class were failed by the old-type part of the Regents examination. The middle curve shows that many students were passed by the old-type who according to the new-type scores are in the lowest 10% of the class in terms of real achievement. Here again the force of the evidence against the old-type physics examination does not depend on the acceptance of the new-type examination as completely valid. That all the new-type tests referred to in this report are valid to a certain extent cannot be doubted by any reasonable person who accepts the validity of the old-type, since the new-type examinations in every case correlate with the old-type as well as, or better than, the old-type results correlate with themselves. When all the evidence, both experimental and *a priori*, is considered, it seems clear that the new-type tests are as much more valid, as they are more reliable, than the old-type; but for the sake of clearness it should be said that the evidence of Charts

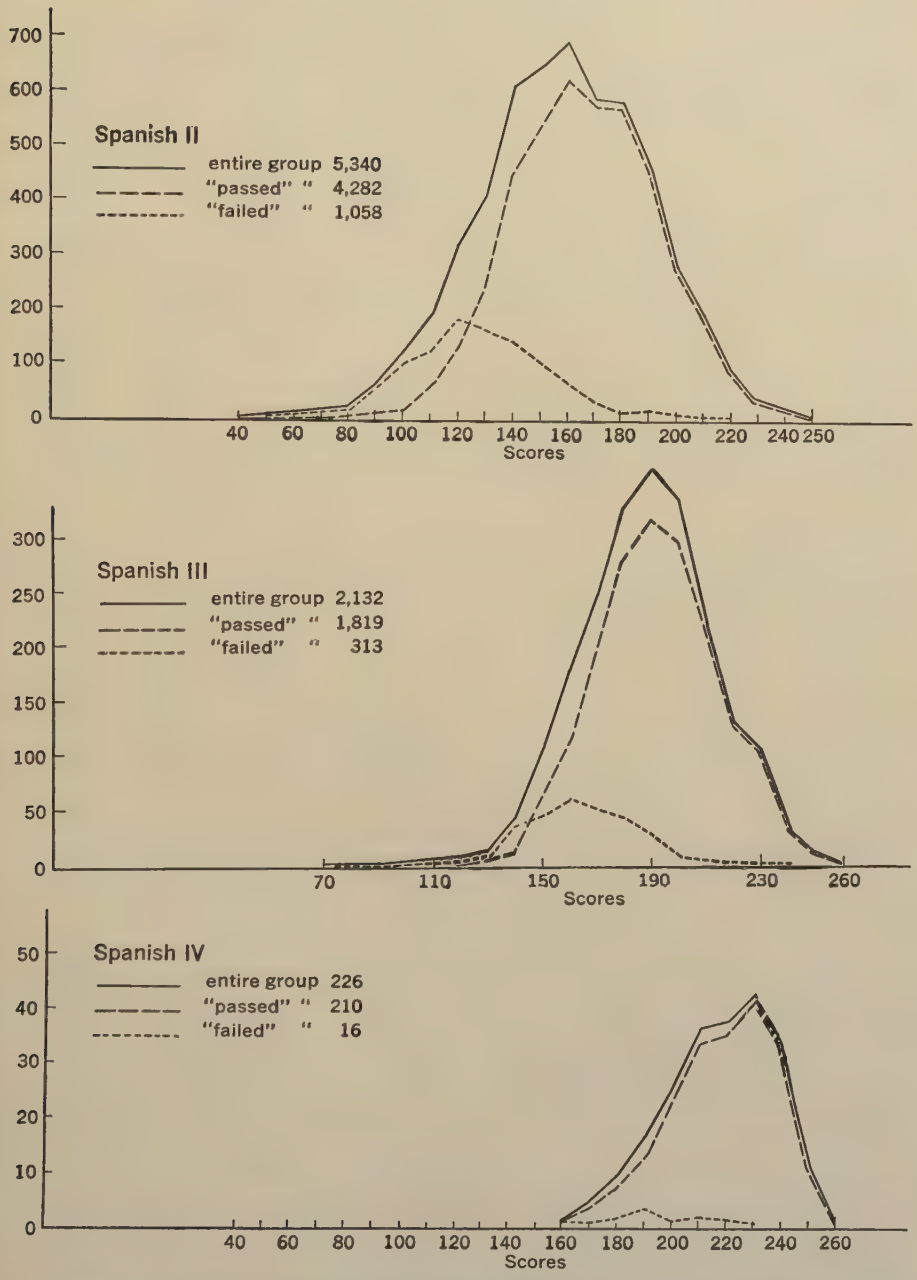


CHART 31. — Graphs of distributions of new-type scores of total population, passed and failed groups of each Spanish class. The passed and failed groups consist of students passed or failed by the old-type parts of the Regents examinations. Scores are represented on the abscissa.

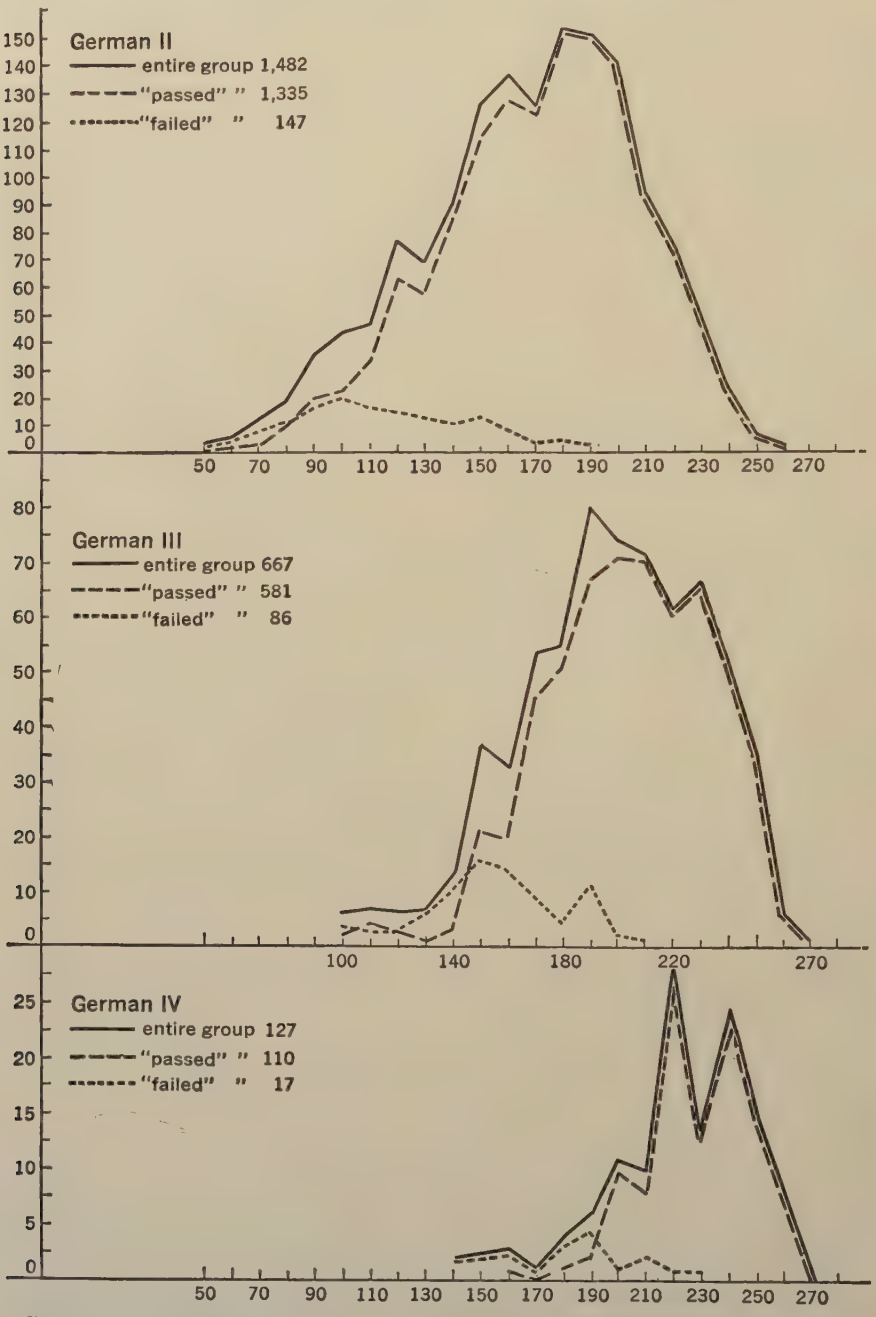


CHART 32. — Graphs of distributions of new-type scores of total population, passed and failed groups of each German class. The passed and failed groups consist of students passed or failed by the old-type parts of the Regents examinations. Scores are represented on the abscissa.

17 to 33 inclusive would be convincing as to the inaccuracy of the old-type examinations even if we knew that the new-type examinations were in each case barely equal in validity and reliability to the old-type examinations.

Forms of distributions are evidence against old-type examinations. — In the preceding charts the reader is able to see that the distributions of new-type scores conform closely to the normal curve. It is generally held by those entitled to an opinion that the educational achievements

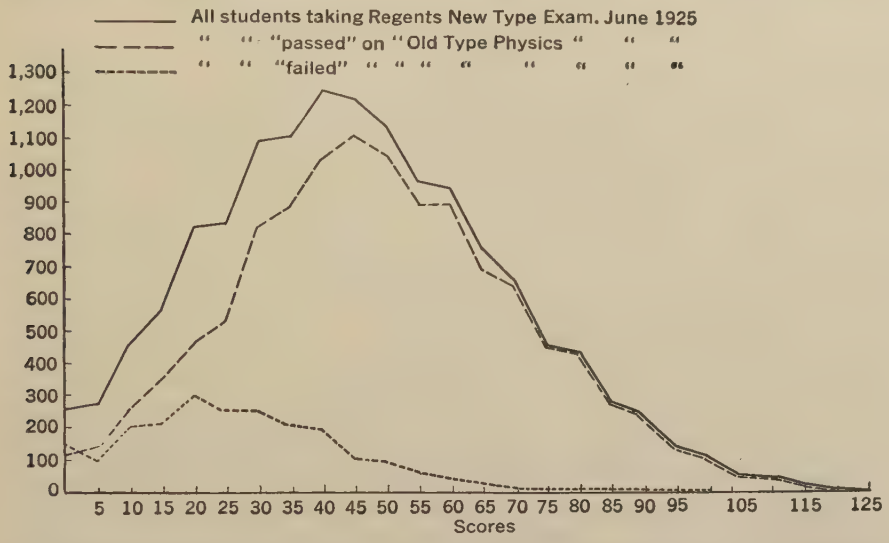


CHART 33. — Graphs of distributions of new-type Physics test scores, of all students taking the Regents physics examination in June, 1925, of students passed and of students failed on the old-type part of the examination.

of students are distributed approximately according to the probability or normal curve. We know from actual experience that most of the physical measurements of school children, such as of height, weight, sitting-height, length of fore-arm, etc., are distributed approximately in the form of the normal probability curve. Accordingly, the form of distribution of scores given by an examining device can be used as evidence of the general validity of the examination. We know, for example, that it is just as contrary to reason to expect that students are divided into two discrete groups with reference to achievement in French, as it is to expect that they are divided into two discrete groups with reference to height. Any yard-stick which would give a U-shaped distribution of the heights of high school juniors would certainly be suspected immediately by all competent teachers.

To divide high school juniors into discrete passed and failed groups is

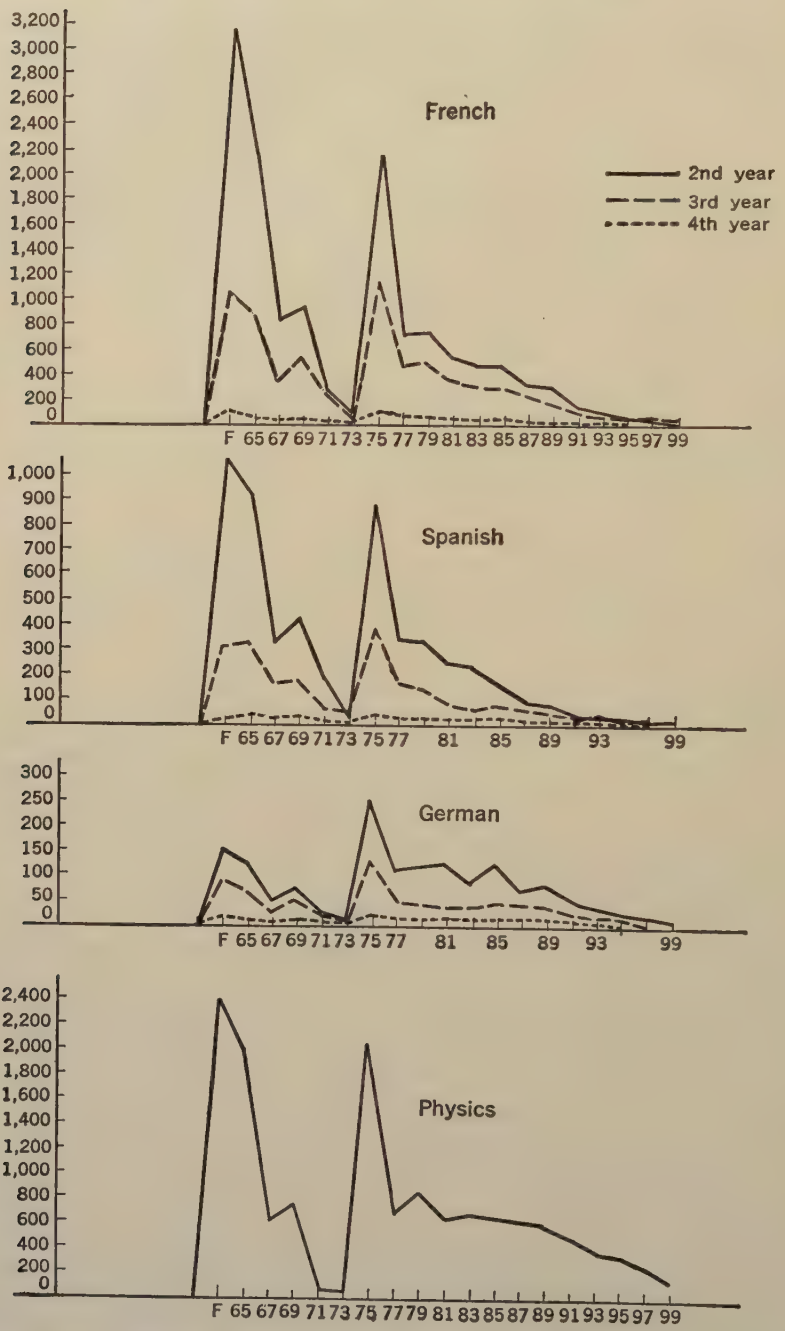


CHART 34. — Graphs of distributions of old-type examination grades by classes. These are graphs of "Regents grades" or of grades from the schools which have been accepted by the Regents reviewers. Percentage grades are represented on the abscissa.

in general terms about as useful to the teachers of the state as dividing the same students into discrete tall and short groups would be to the tailors of the state. The proportion of maladjustments would necessarily be about as great in one case as the other if the true distributions in the two cases are approximately of the same form. Chart 34 shows a very pronounced tendency towards U-shaped curves for all of the old-type examinations in the three modern languages and in physics. Any new-type examination which gave such distributions would immediately be rejected. The shapes of the curves given by the old-type examination grades may be explained in various ways, but such explanations would appear to be only added evidence against the old-type form of examination as the sole criterion of achievement in modern languages and in physics. It is easy to see, for example, that the sharp dip in the curve between percentage grades 70 and 75 is an artificial depreciation which is resorted to on the frank assumption that college admission officers do not like to make close border-line decisions.

Regents examinations should be strengthened, not abandoned or curtailed because of their weaknesses. — Needless to say, the evidence here adduced from the forms of the new- and old-type distributions is not sufficient alone to be very significant one way or another, but in conjunction with all the other evidence presented in the report, it goes to show that the Regents examinations need to be strengthened in order that they may more adequately meet the crucial need for which the Regents examinations were originally instituted. The proposal made by some critics, that the Regents examinations, in view of their weaknesses, ought to be abandoned is given no support whatever by the data of this report. On the contrary, our data show clearly that abandonment or weakening of the Regents or College Entrance Examination Board examination systems would be a retrogression toward complete chaos. The variations and inaccuracies of local school ratings would almost certainly be more anarchical and meaningless than at present if the moral influence of such examining agencies on standards were withdrawn. What we need is to make this influence more accurate and just in its actual operation than it is with the present subjective examinations used by these agencies.

Variable Standards of Schools Not Equalized by Regents Examinations

The standards in individual schools and in groups of schools are, as has already been shown, quite variable. One of the purposes of the Regents examinations is to maintain high and uniform standards of achievement in all subject matters and in all schools in the state. Our data show that the Regents examinations fail to achieve this desirable

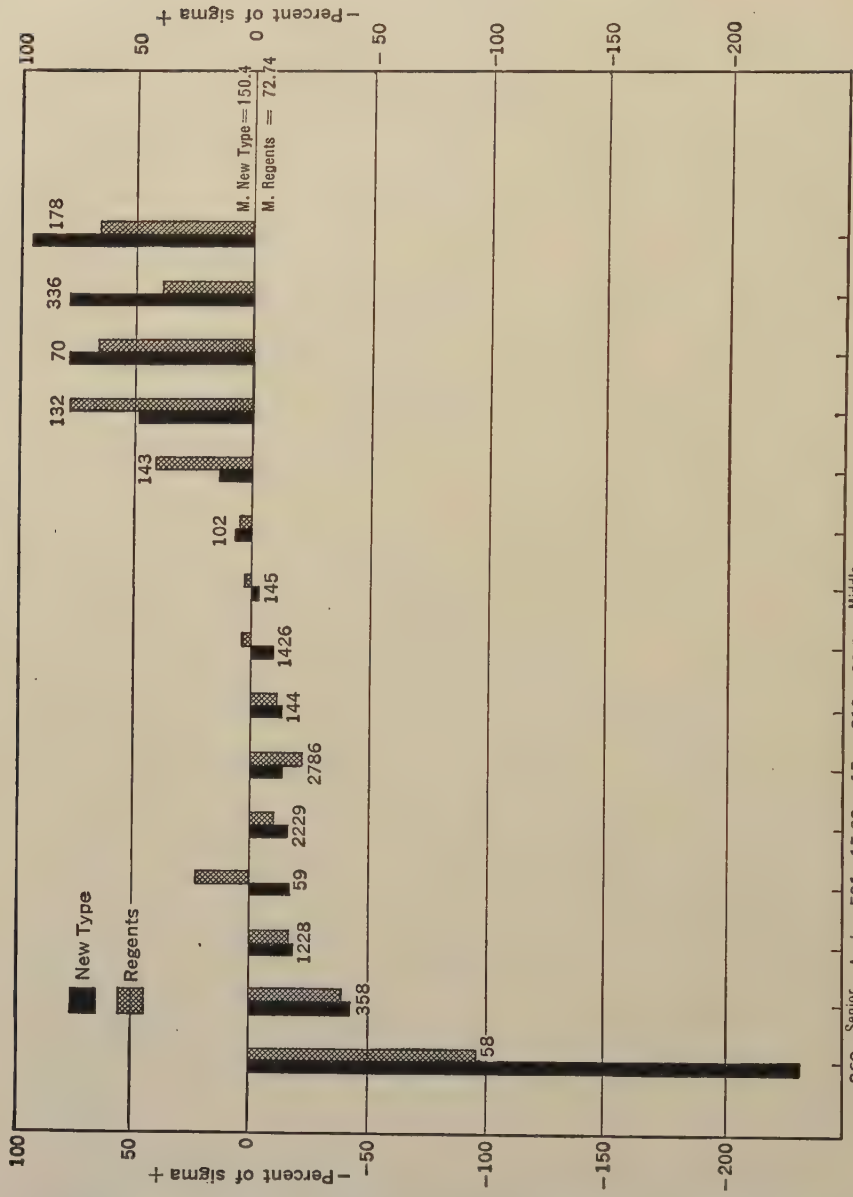


CHART 35. — Variability of Regents standards in French II. The horizontal line from which the vertical bars originate represents the superimposed state-wide means, or averages, of old- and new-type scores. Each pair of bars refers to the students from one school, or a group of schools. (See Table 21, above.) The black bar shows the average new-type score expressed in terms of standard deviation of new-type scores and the hatched bar shows the old-type average expressed in terms of the standard deviation of old-type scores of all French II students. All the bars in the chart are therefore directly comparable with one another. The number of cases is indicated at the end of each pair of bars.

purpose in all of the four subject matters and in each of the modern language classes with which this experiment is concerned. Space limitations do not permit us to display all the instances in which this failure to maintain standards is apparent. We shall therefore take only a meagre sampling and confine our illustrations to second-year French and to physics examination results for several large individual schools and classes of schools.

Variable standards displayed by comparing new- and old-type average scores of French II students in individual schools and groups of schools. — Chart 35 shows for fifteen schools or groups of schools the average scores achieved by students on the new-type French examination and on the French II old-type Regents examination expressed in terms of deviations from the state-wide French II average. The horizontal line from which all the bars in Chart 35 originate represents the state-wide average on both new- and old-type French examinations. The length of each black bar shows the average new-type score of a school or group of schools, expressed in terms of the state-wide standard deviation of new-type scores, and the length of the adjacent cross-hatched bar shows the average score of the same school or group on the old-type French examination, expressed in terms of the state-wide standard deviation of the old-type scores for French II. Therefore all the bars in Chart 35 are directly comparable.

In general there is considerable agreement between the black and cross-hatched bars, but there are several disagreements. Thus school 591, with 59 students in French II, is below average according to the new-type but is somewhat above average according to the old-type examination. Schools 696 and 862 have their positions directly reversed by the new- and by the old-type results. The most striking difference between old- and new-type results shown in Chart 35 is that of school 369, with fifty-eight students. According to new-type results the French II students in this school average more than two standard deviations below the state-wide average. According to old-type scores they average less than one standard deviation below the state-wide average. All these disagreements are almost certainly due to the subjectivity and lack of comparability of the old-type examinations, aggravated by the sampling method of reading to which the Department reviewers in Albany are restricted.

Variable standards displayed by physics test results. — The differences displayed in Chart 36 for new- and old-type physics examination results are even greater than those set forth in Chart 35. Schools 821 and 727 are below average according to new-type and above average according to old-type. School 918 is above average according to new-type and below average according to old-type. School 863 according to new-type is barely above average and according to old-type is almost one-half

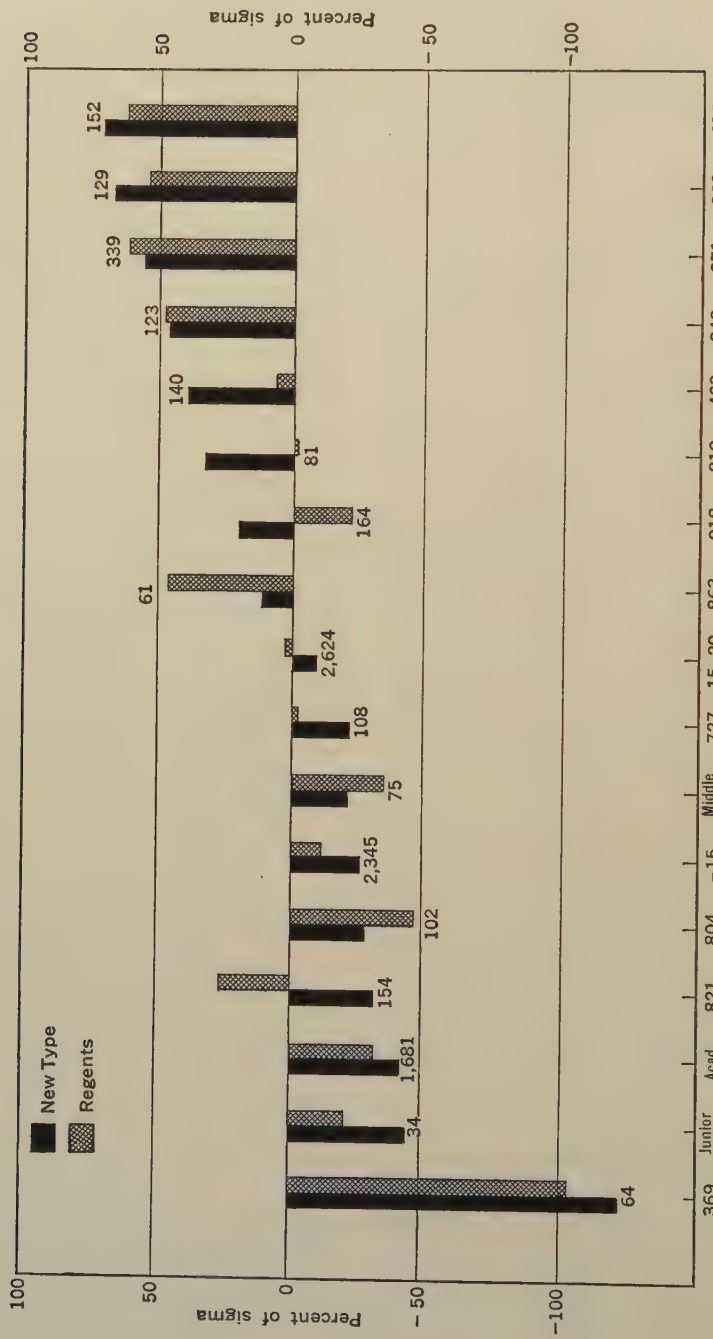


CHART 36. — Variability of Regents standards in physics. Showing for physics results what Chart 35 showed for French II results.

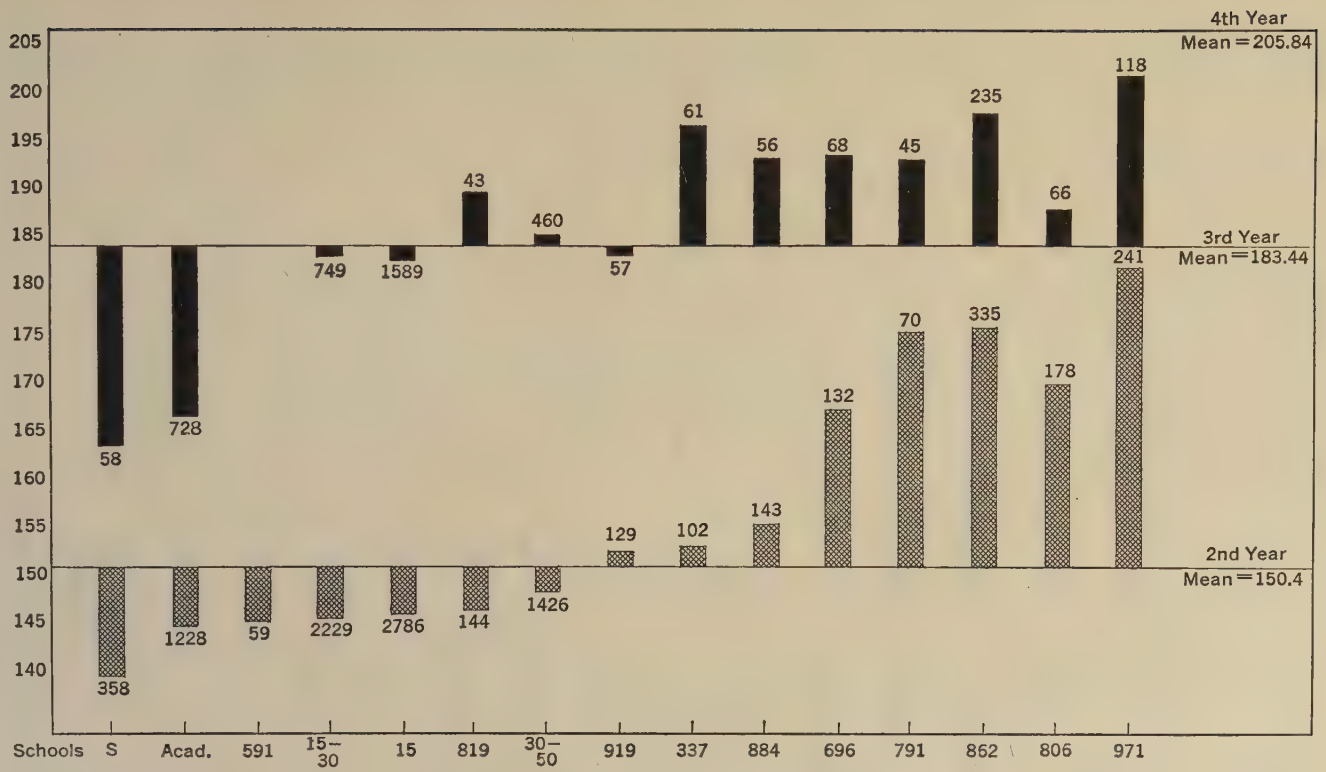


CHART 37. — Graphs of the new-type average scores of all French students taking the Regents examinations in June, 1925, in indicated schools and classes of schools. The horizontal lines from which the vertical bars originate are the second- and third-year class averages on a state-wide basis. The bars are in terms of new-type French test scores, according to the score-scale at the left. The schools and groups of schools involved here are those of Table 21, *q.v.* The number of students is indicated at the end of each bar. Note that large schools, while usually above average, show marked variations.

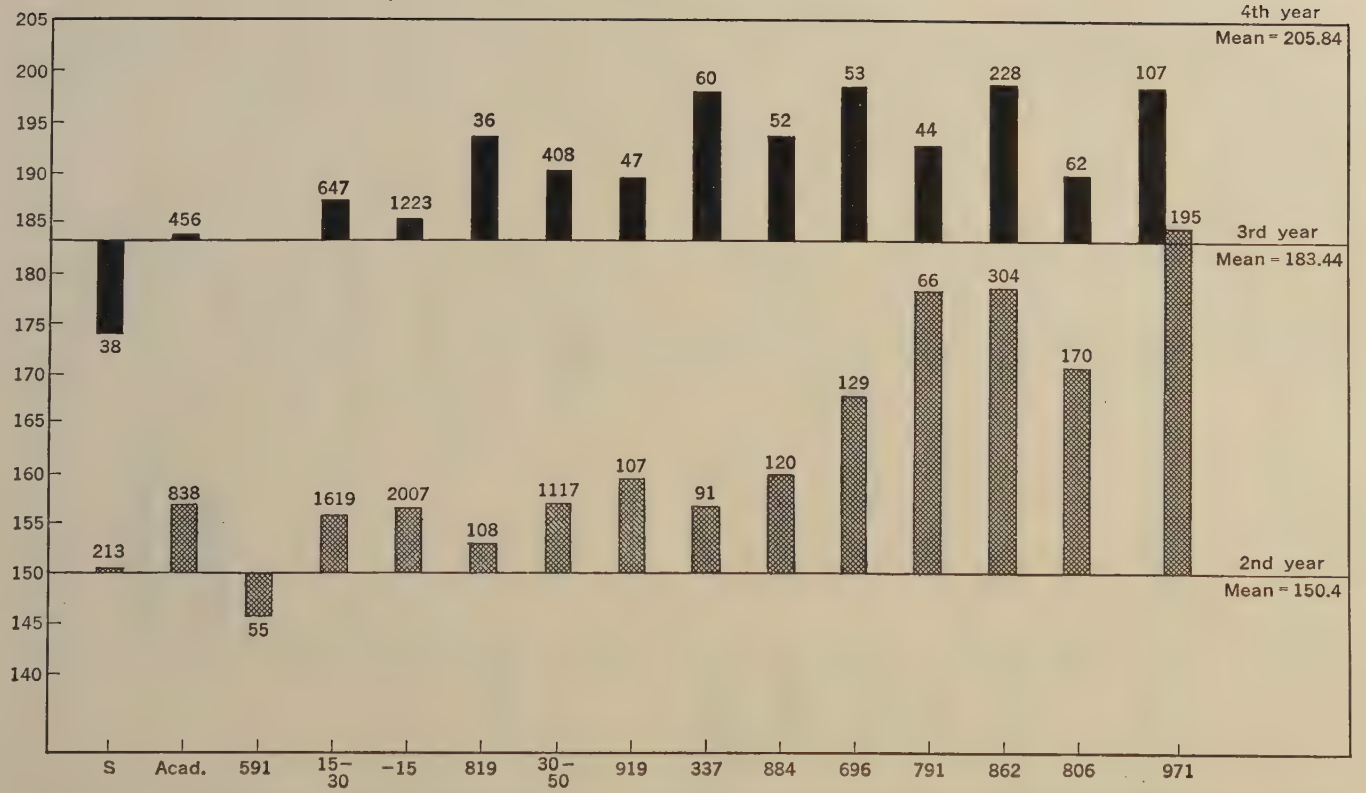


CHART 38. — Showing average new-type scores of French students in individual schools and groups of schools who were either passed by the Department reviewers, or who were allowed to be passed on the basis of school ratings which were reviewed only by the sampling method. This chart parallels Chart 37, and should be read in connection with Charts 37 and 39.

standard deviation above average; while 819 according to new-type is far above average, but according to old-type is a little below average. School 884 is in much the same situation as school 819. These two schools read their physics papers with great severity, and, because of the sampling method of reviewing, their severe reading was accepted by the State Department without mitigation. Schools 821, 727 and 863 read their papers much more leniently and their leniency was not stiffened by the Department reviewers.

In both these charts it is evident that the large schools in general average above the state-wide average, and the smaller schools and academies below average. The lowest school in both French II and physics is No. 369, a private school, while one of the highest schools in both charts is No. 862, which is a large high school in New York City.

Variable standards displayed by comparison of new-type average scores of second- and third-year French classes in individual schools and groups of schools. — The variability of the Regents standards is further displayed in Charts 37, 38, and 39, which show for individual schools and groups of schools the new-type average scores of second- and third-year French students. Chart 37 shows the averages of all the students in these schools that took the Regents French examination in June, 1925. The horizontal lines from which the bars originate represent the second- and third-year state-wide averages. The third horizontal line at the top represents the fourth-year French state-wide average. The length of each bar above or below one of the average lines indicates the average new-type score of all the students in the school or group of schools described at the bottom of the chart. The bars are not in terms of standard deviation but represent raw new-type scores. The presence of the three average lines, however, enables the reader to see at once the extent of the differences between the achievements of students in the same year-class, but in different schools. The figures at the end of each bar show the number of students whose average new-type score the bar represents.

For example, the bar in the lower left-hand corner of the chart shows that the 358 French II students from the senior schools of the state had an average new-type score of approximately 139, which is eleven points below the French II state-wide average; the long bar rising from the second-year average line at the extreme right end of the chart shows that the 241 French II students in school 971 had an average new-type score of about 181, which is 31 points above the state-wide French II average, and only two points below the state-wide French III average. The relative positions of the group of senior schools and of school 971 is practically the same with regard also to French III students.

Chart 38 shows to what extent the Regents examinations took account of the differences displayed in Chart 37. This chart shows the new-type

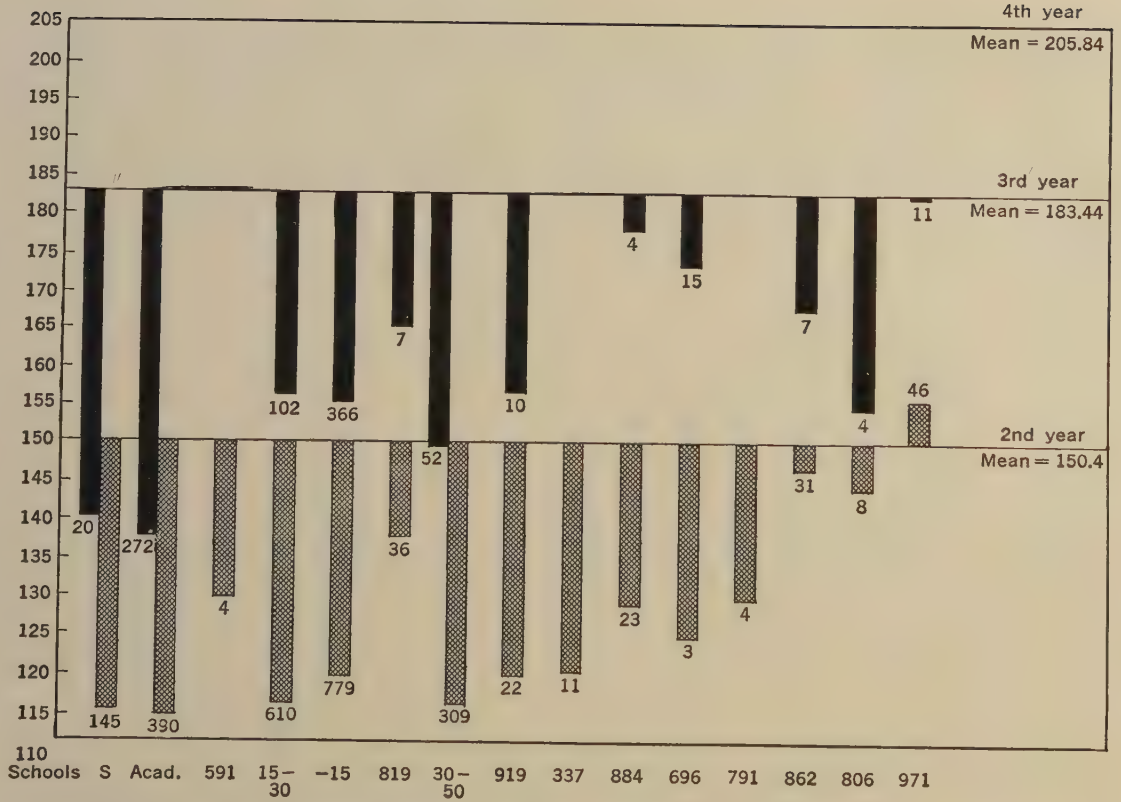


CHART 39. — This chart parallels Charts 37 and 38, and shows the average new-type scores of French students who were failed by the old-type Regents examinations in French II and III.

average scores of students who were passed by the old-type Regents examinations in French II and III. Without attempting to notice all of the interesting features of Chart 38, it is sufficient to observe that the average of the French II students in school 971 who were passed by the old-type Regents examination in French II is actually above the state-wide average for French III and is ten points higher than the average new-type score of the students from the senior schools who were passed in French III. The students passed in French II from school 971 are more than a year above the average of the students passed in French II from school 591. The average new-type scores of the French II passed students from four schools are nearer to the state-wide French III average than to the state-wide French II average. The averages of the French II passed students from three schools are above the average of the French III passed students from the group of senior schools. It is apparent from Chart 33 that the old-type examinations have significantly failed to describe the achievements of students from different schools in terms of uniform standards. Not only this, but the old-type examination in French II has given only French II credit to a group of students who in reality are above the state-wide French III average. This is a notable example of the weakness of a system of separate examinations for each year-class. The fundamental weakness of such a system is that it assumes beforehand, and in contravention of all available evidence, that no student in a given class could possibly deserve to pass in any higher class or to fail in any lower class. Under the Regents rules the rigidity of the separate examination system is somewhat mitigated by special dispensations to a few students, on the recommendation of their teachers, whereby they are allowed to take, for example, the French III examinations when they have had only two years of French classwork. Chart 38 shows clearly the inadequacy of this meagre and reluctant provision for individual differences.

Chart 39 is parallel to Charts 37 and 38 and shows the average new-type scores of French II and III students (from the same schools and groups of schools) who were failed by the old-type Regents examinations. The average new-type scores of students from senior schools and academies who were failed in French III are far below the state-wide average of French II students, while the students failed in French II from schools 884, 696, and 971 secured average new-type scores almost equal to the new-type French III average. In other words, the students who were failed in French III from the senior schools and academies are more than a year below those failed from some other schools. The students from the senior schools and academies who were failed in French II are considerably more than a year below the average of the students failed in French II from school 971. The failing standard thus varies even more

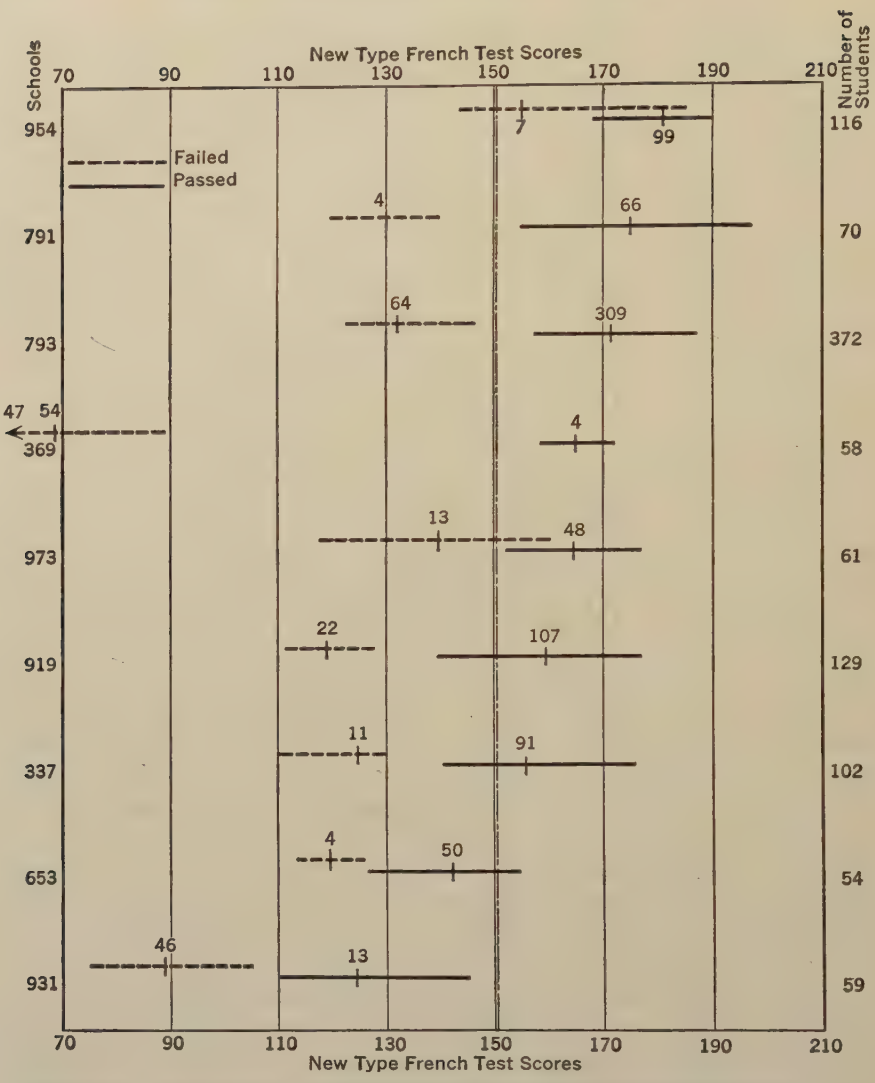


CHART 40. — Showing for indicated schools the medians and interquartile ranges of new-type scores in French II. The school identifications appear at the left, the numbers of passed and of failed students are indicated just above the interquartile range lines, and the numbers of both passed and failed students are indicated at the right. The vertical dash-dot line indicates the state-wide average of all French II students, both passed and failed.

than the passing standard. In school 971 eleven students were failed in French III who are almost exactly at the French III state-wide average and forty-six students were failed in French II who averaged five points above the state-wide French II average. The average of the stu-

dents failed in French II in schools 862, 806 and 971 are above the averages of the students failed in French III from the senior schools and academies.

The variability of standards displayed in these three charts is in reality greater than appears because in these charts we are dealing only with averages, and not with extreme individual variations.

Looking again at Chart 37 one would expect that a much larger proportion of the students at the left of the chart would be failed than of the students from the schools at the right end of the chart. Comparing the numbers at the ends of the bars in Chart 37 with the numbers at the end of the corresponding bars in Chart 39, we may ascertain to what extent the Regents examinations fulfilled this expectation. About 40% and 32% of the French II students from the two lowest school groups in Chart 37 were failed by the old-type part of the Regents examination; the corresponding proportions for the four schools in the middle of the chart ranged from 10 to 22%; but of the students from school 971, 19% were failed in French II. In other words, almost a fifth of the French II class in school 971, which in reality was barely below the French III state-wide average, were failed by the old-type part of the Regents examination, while considerably less than a fifth of the French II students from schools 919, 337 and 884 were failed. Such facts as these hardly need comment.

Variability of Regents standards only one aspect of fundamental inaccuracy of old-type subjective examinations. — The variability of the Regents standards is partly due to the fact that the State Department of Education accepts a part of the school ratings without review. That is, many school ratings are accepted by the Department readers on the basis of the review of a relatively small sample of school ratings. The standards of different readers without central supervision naturally vary considerably, but that there is also a considerable element of inaccuracy behind the variations displayed in Charts 37 to 39 is indicated in Charts 40 to 42, which show the interquartile ranges of the passed and failed students from several different schools.

Chart 40 shows that the middle 50% of the students *failed* in school 954 are entirely above the middle 50% of the students *passed* in school 931. More than a quarter of the students *failed* in school 954 are above the average of those *passed* in that school by the old-type examination. In school 973 more than a quarter of the students failed achieved higher scores than the lower third of those that passed. Nearly three-quarters of the students failed in 954 are above the state-wide average and about a third of the students passed in school 931 are below the lower quartile of all French II students who took the Regents examinations in June, 1925.

Charts 41 and 42 are so similar to Chart 40 both in form and indications that they are presented without comment.

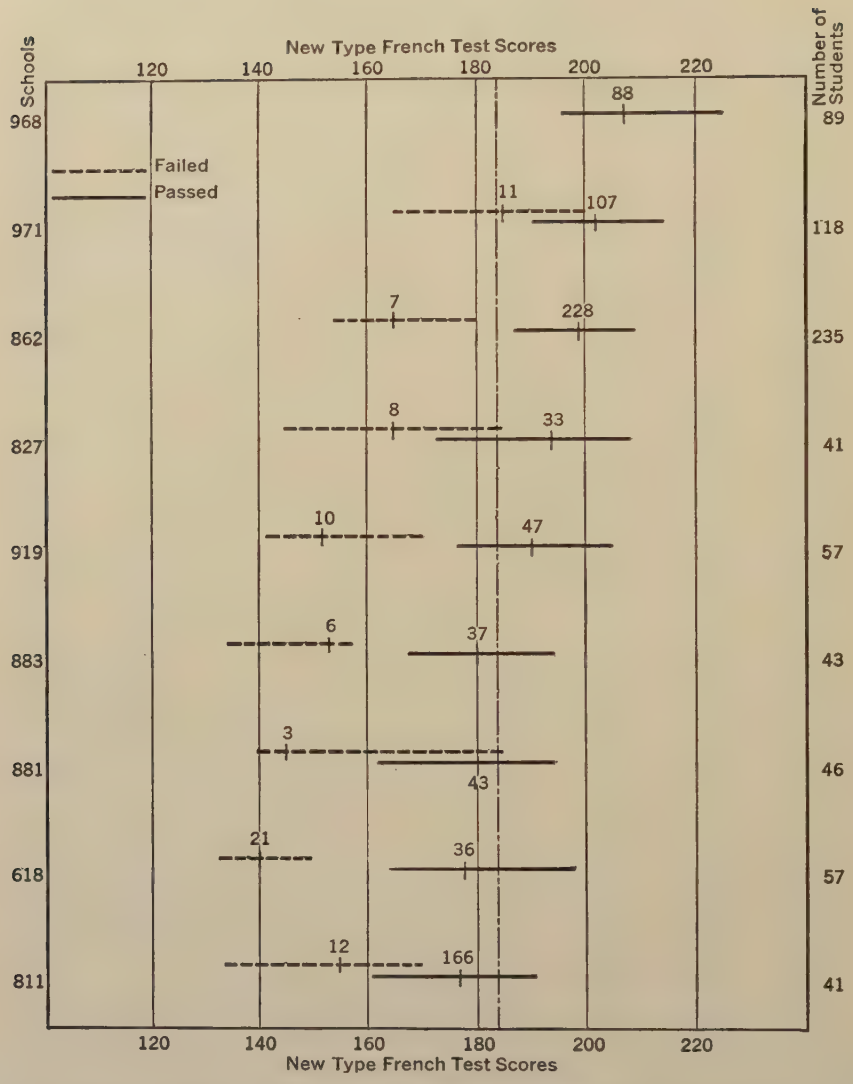


CHART 41. — Showing for indicated schools the medians and interquartile ranges of new-type scores in French III. The school identifications appear at the left, the numbers of passed and of failed students are indicated just above the interquartile range lines, and the numbers of both passed and failed students are indicated at the right. The vertical dash-dot line indicates the state-wide average of all French III students, both passed and failed.

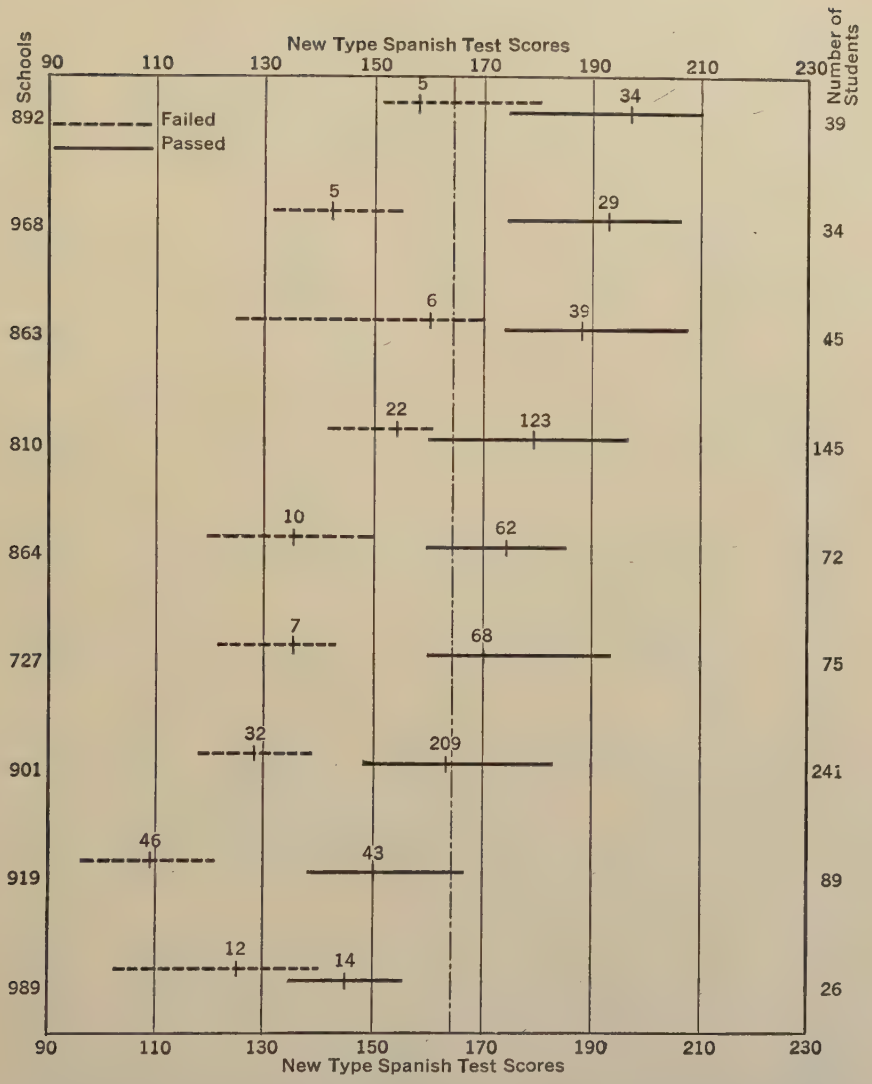


CHART 42. — Showing for indicated schools the medians and interquartile ranges of new-type scores in Spanish II. The school identifications appear at the left, the numbers of passed and of failed students are indicated just above the interquartile range lines, and the numbers of both passed and failed students are indicated at the right. The vertical dash-dot line indicates the state-wide average of all Spanish II students, both passed and failed.

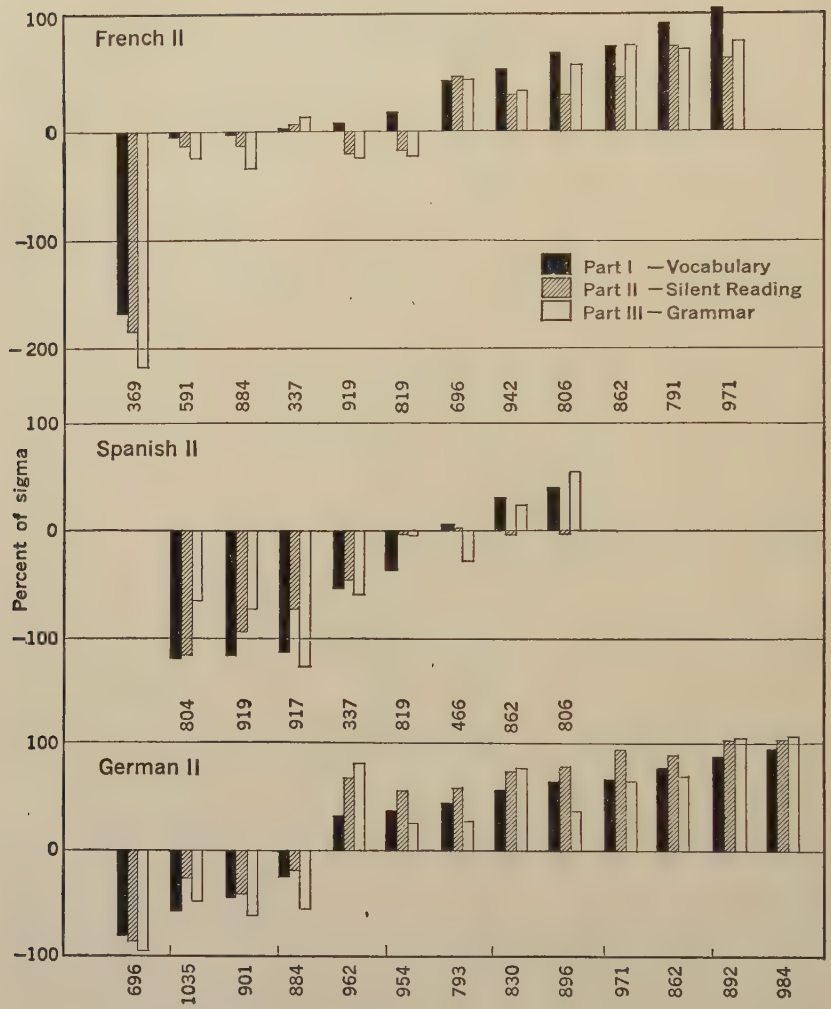


CHART 43. — Showing "unevenness" of achievement in modern language classes with respect to vocabulary, silent reading, and grammar. Only second-year classes are considered because of small numbers of students in third- and fourth-year classes. The bars are in terms of state-wide standard deviations, and the horizontal lines from which they originate represent superimposed state-wide means of the three parts of the new-type examinations. Each triad of bars represents one school, and the bars in each triad represent the parts of the new-type tests in order, from left to right, vocabulary, silent reading, and grammar.

Unevenness of achievement in individual schools with respect to vocabulary, reading and grammar. — From the correlations between the Parts of the new-type tests presented in earlier sections of this report, it is clear that achievement of individual students is very uneven with respect to vocabulary, reading and grammar. That is, a student may rank very high with respect to grammar and very low with respect to vocabulary, or he may rank very high in reading ability and very low in grammar. It is of interest to know whether similar differences exist as between different schools. Is grammar, for example, emphasized relatively more in one school than in another? Are certain schools high in reading and low in grammar achievement? Chart 43 has been prepared with a view to answering such questions.

The horizontal lines from which the vertical bars originate represent the superimposed state-wide means of the three Parts of the new-type examinations in French, Spanish and German. Each triplet of bars represents one school and the bars in each triplet represent the Parts of the new-type tests in order. The length of the bar above or below the mean indicates the average of the school on the given part, expressed in terms of the state-wide standard deviation of scores on that part of the new-type test. All the bars originating from the same horizontal line are therefore directly comparable. While there are no very great differences, there are several which seem significant. Compare, for example, schools 884 and 337 in French II. These two schools are almost at par with respect to vocabulary but school 884 is a little over three-tenths of a standard deviation below and school 337 is .15 of a standard deviation above the state-wide average in grammar. Only one school is equal in all three parts, that is school 696. Only two of the schools chosen for Chart 43 are higher in grammar than in any other part, schools 337 and 862.

The differences for Spanish II are somewhat greater than those for French II. School 804 is considerably better in grammar than in either vocabulary or reading; while school 917 is very poor in both grammar and vocabulary but less poor in reading. Schools 862 and 806 are at par with respect to reading but considerably above average in grammar and vocabulary. These two schools are about equal to 819 and 466 with respect to reading but very different with respect to grammar and vocabulary.

With respect to German II, schools 884 and 962 show very different achievements in grammar. Both these schools are much nearer to the state-wide average with respect to vocabulary and reading than with respect to grammar.

On the whole it appears that some schools do emphasize grammar more than others, while other schools emphasize either vocabulary or reading.

V

INTERNAL STRUCTURE OF OLD- AND NEW-TYPE MODERN LANGUAGE EXAMINATIONS

Thus far we have presented experimental evidence on the reliability, validity and comparability of the old- and new-type modern language examinations. The demonstrated advantages of the new-type over the old-type are in part due to the fact that only a sampling of the old-type papers are actually read by the central reviewing authorities, and to other unfortunate features of the Regents system that are not necessarily inherent in the old-type subjective form of examination. We shall now attempt to show that, even if these unnecessary disadvantages under which the old-type operates in the Regents system were removed, the inherent and unavoidable limitations of the subjective form of examination are so strict that it probably could never, under practical working conditions, equal the new-type in reliability, validity, comparability or administrative convenience, hour for hour of examination time, and dollar for dollar of cost.

In order to justify this statement we shall consider in some detail the method of construction of the old- and the new-type tests used in this experiment, the size and character of the samplings of vocabulary elements, and the nature and measurement values of the various activities demanded of the student in the two types of examinations. In order to make our comparison as valid as possible, we shall include in our analysis: (*a*) one new-type French examination not used in the Regents experiment; (*b*) the 1924 as well as the 1925 old-type Regents examinations in French, and (*c*) the 1924 and 1925 College Entrance Examination Board French examinations.

Methods of construction. — The old-type modern language examinations used in this experiment were constructed and edited in accordance with the subjective judgments of committees composed of notable scholars and teachers. The members of the committees were appointed long before the examinations were actually made and some members had served on similar examination committees for several years, and therefore had considerable experience back of their judgment in such matters. There was one committee for each language, the State Supervisor of Modern Languages being an ex-officio member of all three committees. The examination committee for each language constructed the second-, third- and fourth-year examinations in that language. Each member of

each committee brought to the meeting at which the final form of the examination was decided upon, carefully prepared questions for one or several parts of the examination. The offerings of each member of the committee were discussed by the committee as a whole and the questions finally adopted represented the best questions brought to the meeting by the various members. There is no means of knowing how much time the members of the committees spent on their tasks before they came to the one and only conference of examination makers. But the conference at which one set of second-, third- and fourth-year examinations were finally constructed took up the better part of a working day. The writer of this report was privileged to meet with one or two of these committees and was greatly impressed by the seriousness with which the members took their tasks, and with the care exercised in criticizing and in improving the questions that were finally adopted for each of the examinations. The most impressive feature of the conference, however, was that all questions regarding the relative difficulties of two or more words, or of two or more sentences or paragraphs, or regarding the relative measurement values of two similar or dissimilar questions, were decided, as far as could be made out, entirely by subjective judgment. There was no reference made at any time during the meetings to any *objective evidence* concerning the difficulty or the measurement value of any of the questions discussed. There were frequent disagreements on such questions, and after free, and often spirited, discussion or argument the matter was settled by a vote of the members of the committee. After the examinations had been made by these committees they were taken to Albany by the supervisor and there passed through several editorial processes.

Reproductions of old-type examinations. — To illustrate the general character of the examinations constructed in this manner, the Regents French and physics papers used in June, 1925, are reproduced here. The reader is invited in passing to notice particularly two features of these old-type examinations. (*a*) The directions state that "The minimum time requirements (*for admission to the examination*) is five recitations a week for" a certain number of years (see second paragraph of heading of each examination below, pp. 200, 202, 204 and 206). This is the same concrete manifestation of the old time-serving conception of education as is found on the College Entrance Examination Board examinations. It is a curious survival in an age which boasts the discovery and constructive exploitation of individual differences. (*b*) Modern language students are advised that "credit for oral work," as determined by local standards, may be substituted for question 7, which calls for an original composition in the foreign language (see last paragraph of *Note to the Student*, pp. 200, 202 and 204). These two features will be discussed in later sections of this report (pp. 306 and 317).

The University of the State of New York

233D HIGH SCHOOL EXAMINATION

FRENCH — Two Years

Thursday, June 18, 1925 — 1.15 to 2.45 P.M., only

Write at top of first page of answer paper (*a*) name of school where you have studied, (*b*) number of weeks and recitations a week in French 1 and 2.

The minimum time requirement is five recitations a week for a school year in (*a*) first year French, (*b*) second year French.

NOTE TO THE STUDENT

This part of the examination is limited in time to an hour and a half. Do not waste time. Do not make a preliminary draft to be copied later. Do first whatever question you can do most readily, but number your answers to correspond to the questions.

Students who are entitled to credit for oral work may omit question 7. Students who are not entitled to credit for oral work should answer all the questions. Credit for oral work may be added on the answer paper directly to the mark obtained in the written examination.

1 Traduire en anglais: (Two credits to a line) [25]

UN JOUR DE VACANCES

C'était un jour de vacances: grande fête pour les enfants qui n'aimaient rien tant que d'aller à l'école, si ce n'est (*unless it be*) de rester à la maison et d'avoir devant eux toute une journée pour faire ce qui leur passerait par la tête. La mère était joyeuse aussi de garder ses enfants autour d'elle, mais pourtant un peu inquiète et effrayée d'avoir à gouverner pendant douze heures, sans aucun secours (*help*), tout ce petit peuple tapageur (*noisy*), toujours entre le rire et les larmes, entre une joie bruyante (*noisy*) et des désespoirs plus bruyants encore. Elle s'était endormie la veille au bruit de la pluie qui tombait, tombait, une de ces pluies tranquilles et abondantes qui semblent avoir une ferme résolution de ne jamais s'arrêter. En se réveillant, elle l'entendit encore. C'était un peu décourageant.

2 Traduire en français: [20]

When we awoke this morning at half past seven it was raining. We couldn't play in the yard (garden) because of the rain. After breakfast mother wished us to sit in the living room and read but George preferred to play cards. We had just sat down when my uncle came in and we had to listen to his stories. In the afternoon the rain stopped and we went for a walk. So we had a good time after all.

3 Écrire les temps primitifs et la première personne singulier du futur de *trois* des verbes suivants: lire, écrire, croire, sortir, voir, prendre. [15]

4 Compléter *cinq* des phrases suivantes: [5]

a Il m'aide . . . faire mes devoirs.
b Apprenez-vous . . . parler français?
c Ils n'ont plus . . . argent.
d . . . quoi pensez-vous?
e Nous essayons . . . trouver ces papiers.
f J'ai . . . papier jaune.
g C'est une femme . . . je connais le fils.

5 Mettre la forme convenable du verbe entre parenthèses: [5]

a Nous désirons qu'il (*venir*).
b C'est moi qui les (*avoir*) écrits.
c Elle s'est (*casser*) le doigt.
d Nous ne nous sommes pas (*voir*).
e Je ne crois pas qu'elle (*devoir*) faire cela.

6 Traduire en français *cinq* des expressions suivantes: [10]

(*a*) You ought to do it, (*b*) four months ago, (*c*) while reading his lesson, (*d*) He must work, (*e*) It is late, (*f*) He is late, (*g*) They should have looked for it.

7 Écrire en français une composition de 75 mots environ sur *un* des sujets suivants: [20]

(*a*) La France, (*b*) Les Français, (*c*) La langue française (le français), (*d*) L'école (le bâtiment, les matières, les professeurs).

The University of the State of New York

233D HIGH SCHOOL EXAMINATION

FRENCH — Three Years

Thursday, June 18, 1925 — 1.15 to 2.45 P.M., only

Write at top of first page of answer paper (*a*) name of school where you have studied, (*b*) number of weeks and recitations a week in French 1, 2, and 3.

The minimum time requirement is five recitations a week for a school year in (*a*) French 1, (*b*) French 2, (*c*) French 3.

NOTE TO THE STUDENT

This part of the examination is limited in time to an hour and a half. Do not waste time. Do not make a preliminary draft to be copied later. Do first whatever question you can do most readily, but number your answers to correspond to the questions.

Students who are entitled to credit for oral work may omit questions 6 and 7. Students who are not entitled to credit for oral work should answer all the questions. Credit for oral work may be added on the answer paper directly to the mark obtained in the written examination.

1 Traduire en anglais: (Two credits to a line) [25]

LE NOVEAU PENSIONNAIRE

La femme et son fils étaient venus tous les deux en voiture. Elle était veuve, et fort riche, à ce qu'elle nous fit comprendre. Elle avait perdu le cadet de ses deux enfants, qui était mort un soir au retour de l'école, pour s'être baigné avec son frère dans un étang malsain. Elle avait décidé de mettre l'aîné, Augustin, en pension chez nous, pour qu'il pût suivre le Cours Supérieur. Et aussitôt elle fit l'éloge de ce pensionnaire qu'elle nous amenait. Ce qu'elle contait de son fils avec admiration était fort surprenant: il aimait à lui faire plaisir, et parfois il suivait le bord de la rivière, jambes nues, pendant des kilomètres, pour lui rapporter des œufs de canards sauvages. L'autre nuit il avait découvert dans le bois une faisane prise au collet dans des nasses qu'il avait tendues.

2 Traduire en français: [20]

a Do you intend to write to your mother today?
b If he has enough money he will give us some.
c Did you have a good time in the country last week?
d He succeeded at last in buying the house.
e I wish that you would speak more distinctly.
f After leaving the house I lost my tickets.
g She went home at three o'clock.

h How long have you been studying French?

i I am glad I live so near the school.

j It is very warm now, every day, even when it rains.

3 Écrire les temps primitifs des verbes suivants: voir, lever, devenir, sentir, pleuvoir. [10]

4 Traduire en français: (*a*) Give it to him, (*b*) Do not tell it to her, (*c*) What are you thinking of? (*d*) What do you think of him? (*e*) What fell? [10]

5 Employer *cinq* des expressions suivantes dans des phrases courtes et traduire les phrases en anglais: à la bonne heure, de bonne heure, à l'heure, à temps, tout de suite, tout à coup, au devant de, à la rencontre de. [10]

6 Compléter les phrases suivantes: [5]

(*a*) Ma mère désire que . . . , (*b*) Je le ferais, si vous . . . , (*c*) Il faut que l'élève . . . , (*d*) Il est vrai que . . . , (*e*) Qui que vous . . . , voici votre maître.

7 Écrire en français une composition de cent mots environ sur *un* des sujets suivants: (*a*) Un livre français, (*b*) Un voyage, (*c*) Un Français célèbre. [20]

The University of the State of New York

233D HIGH SCHOOL EXAMINATION

FRENCH — Four Years

Thursday, June 18, 1925 — 1.15 to 2.45 P.M., only

Write at top of first page of answer paper (*a*) name of school where you have studied, (*b*) number of weeks and recitations a week in French 1, 2, 3 and 4.

The minimum time requirement is five recitations a week for a school year in (*a*) French 1, (*b*) French 2, (*c*) French 3, (*d*) French 4.

NOTE TO THE STUDENT

This part of the examination is limited in time to an hour and a half. Do not waste time. Do not make a preliminary draft to be copied later. Do first whatever question you can do most readily, but number your answers to correspond to the questions.

Students who are entitled to credit for oral work may omit question 4. Students who are not entitled to credit for oral work should answer all the questions. Credit for oral work may be added on the answer paper directly to the mark obtained in the written examination.

1 Traduire en anglais: (Two credits to a line) [25]

JEANNE D'ARC

Une enfant de douze ans, une toute jeune fille, confondant la voix de son cœur avec la voix du ciel, conçoit l'idée étrange, improbable, absurde, si l'on veut, d'exécuter la chose que les hommes ne peuvent plus faire, de sauver son pays. Elle couve cette idée pendant six ans sans la confier à personne; elle n'en dit rien, même à sa mère, rien à nul confesseur. Sans nul appui de prêtre ou de parents, elle marche tout ce temps avec Dieu dans la solitude de son grand dessein. Elle attend qu'elle ait dix-huit ans, et alors immuable elle l'exécute malgré les siens et malgré tout le monde. Elle traverse la France ravagée et déserte, les routes infestées de brigands, elle s'impose à la cour de Charles VII, se jette dans la guerre et dans les camps qu'elle n'a jamais vus.

2 Traduire en français: [25]

Two men were neighbors and each of them had a wife and several little children. One of the men was uneasy, saying, "If I die or if I fall ill, what will become of my wife and children?" This thought never left him and it gnawed at his heart as a worm gnaws the fruit in which it is hidden. But the other man lived in peace. "Because," said he, "God, who knows all his creatures and watches over them, will also protect me and my wife and my children."

3 Écrire en français une composition de 100 mots environ sur un livre que vous avez lu *ou* un voyage que vous avez fait. [20]

4 Répondre à toutes les demandes suivantes:

a Indicate the main difficulty or peculiarity in the sound of *one* consonant in *each* of the following words or groups of words: absurde, soixante, second, calme, Bruxelles, nom anglais, cent un, vaciller, sang impur, grand homme. [10]

b Write *five* French sentences, using in each a different compound conjunction followed by the subjunctive. [10]

c Translate: by going; without going; instead of going; after returning; I have finished reading. [10]

The University of the State of New York

233D HIGH SCHOOL EXAMINATION

PHYSICS

Thursday, June 18, 1925 — 9.15 to 10.45 A.M., only

Write at top of first page of answer paper (*a*) name of school where you have studied, (*b*) number of weeks and recitations a week in physics, with the total number of laboratory periods and the length of such periods.

The minimum time requirement is five recitations a week for a school year. A double laboratory period counts in place of one recitation. Laboratory work equivalent to 30 double periods is required.

This part of the examination is limited in time to an hour and a half.

Answer five questions, selecting at least one from each group.

Group I

Answer at least one question from this group.

1 Using a single sentence in each case, state (*a*) the effect of change of temperature on the speed of sound [2], (*b*) the cause of differences in the intensity of sound [2], (*c*) how the particles of air move when a sound wave is coming toward one [2], (*d*) how an echo is produced [2], (*e*) how the length of an organ pipe affects the pitch of the sound produced [2].

2 Name the chief method of heat transference in *each* of the following cases: (*a*) from the sun to the earth [2], (*b*) through the iron of a stove or of a radiator [2], (*c*) in heating a room with a hot air furnace [2]. In *one* of the above cases describe the process by which the heat is transmitted [4].

3 Describe an experiment to show how the principal focus of a convex lens is found [5]. If the focal length of the lens is found to be 16 cm, how far from the lens is the image of an object placed 32 cm from the lens [5]?

Group II

Answer at least one question from this group.

4 On a 110 volt circuit, find (*a*) the cost of operating for two hours, at 8 cents a kilowatt hour, a projection lantern equipped with a 1000 watt lamp [4], (*b*) the current flowing through the lamp [3], (*c*) the resistance of the lamp [3].

5 By means of a labeled diagram show the path of the current through a continuous ringing electric bell [5]. Add to the diagram so as to show a push button and two dry cells connected as in a house [3]. Are the parts of this circuit connected in series or in parallel [2]?

Group III

Answer at least one question from this group.

6 State the principle of Archimedes [3]. Explain why it does or does not apply to a floating body [4]. State just why the following statement is incorrect: "Ice formed in sea water is less dense than that formed in fresh water, since sea water ice floats higher than fresh water ice" [3].

7 A loaded sleigh weighing 3000 pounds is just prevented from sliding down a hill that rises 8 feet in 100 feet of road. What is the force exerted by the team of horses [5]? How much work against gravity would the team do in drawing the sleigh 200 feet up the hill at uniform speed [5]?

8 State the law of conservation of energy [3]. Give an example of a body possessing kinetic energy and one of a body possessing potential energy [2]. State the law of motion illustrated when a gun kicks [2]. Explain its application to this case [3].

Method of constructing new-type tests. — The construction of the new-type modern language examinations used in this experiment was a much longer process. The forms of the questions were decided upon after three years of extensive experimentation with more than 20 different forms of modern language examination questions. The vocabulary samplings used were based upon word counts of sixteen widely used textbooks in two languages and upon a synthesis of various word counts of textbooks and syllabi in the third language. The grammar and idiom content of the new-type tests was determined item by item, on the basis of extensive syntheses of standard grammars and syllabi. These selections of materials were finally checked up by actual experiment. No item was accepted for final use unless it gave a positive correlation with known criteria of achievement. Each new-type examination was constructed to cover the whole range of achievement in high school modern language work. That is, each examination included some items easy enough to be answered correctly by the poorest students at the end of the first year and some questions so hard that they could be answered only by the best students at the end of the fourth year. The "easiness" or "hardness" of each individual item was determined upon the basis of the responses by numbers of modern language students in high schools ranging from several hundred to several thousand. No appeal was made at any time to the subjective judgment of scholars regarding the difficulty or the measurement value of an item.

The most striking difference between the two methods of construction, aside from the difference in quality of the resulting products, is that the new-type method shows a much more adequate appreciation than the old-type method of both the difficulty and the importance of constructing

sound examinations. In the old-type very much is left to subjective opinion and to chance; in the new-type everything that is capable of objective verification is experimentally verified. It is because of the reliance on experimental evidence that the new-type examinations are constructed so as to cover the whole range of achievement in a single examination, thus making possible comparability between the achievements of the various classes; and that such comprehensive examinations are at the outset made in several equivalent forms, thus making possible comparability of measurements at different stages of the student's progress.

Advantages of one *examination for all year-classes.* — The advantages of *one* examination covering the *whole range* of achievement in high school modern language work has been demonstrated in the comparisons made in previous sections of this report between the achievements of second-, third- and fourth-year classes. It is obvious that such comparisons could not be made on the basis of the old-type examinations. The value of using forms of examinations which are known to be equivalent in difficulty and in representative character of modern language materials, may be illustrated by comparisons which we are enabled to make between the vocabulary achievements of students of Spanish in the junior high schools of New York City and in the senior high schools in New York State. The 100-word Spanish vocabulary test used in the new-type Regents examination happened to be exactly equivalent to that in the new-type test used in the junior high schools of New York City in June, 1925. The median scores of the 9B and RD junior high school classes on the Spanish vocabulary test were 57 and 58, respectively.[1] The state-wide medians of Spanish II and III students on the vocabulary Part of the new-type Regents examination are 70 and 81, respectively. It is quite reasonable to infer that Spanish I senior high school students would achieve an average score at the end of the course, i.e., at the end of one year of study, about 12 points below the average of Spanish II students, i.e., about 57 or 58. Thus we have objective evidence in support of the theory that the junior high school students on the average achieve about as much in four or three semesters as senior high school students (who survive to the end of the second year) achieve in two semesters. The overlapping of the junior high school fourth- and third-semester students and of the second-year senior high school students is about the same as the overlapping of the latter and the third-year senior high school students. About 13% of the 9B and RD Spanish students in the junior high schools surpass the average of the Spanish II senior high school students with respect to vocabulary. About one-sixth of the "rapid advancement" students in the junior high school Spanish classes achieve as much or

[1] The 9B class is the fourth-semester class of "normal" students and the RD class is the third and final class of "rapid advancement" students.

more in three semesters with respect to vocabulary as the average Spanish II students after four semesters in senior high school Spanish work.[1]

Reproductions of new-type Regents examinations of June, 1925. — Except for the columns of figures at the left, and except for minor differences in size of type and in spacing, the new-type parts of the Regents examinations of June, 1925, are reproduced below exactly as they were presented to the high school students. Correction of some obvious defects has been omitted so that the reader may secure as accurate a picture of the experiment as possible. The first column of figures at the left shows for each of the questions the per cent of random samplings of students that answered it correctly. Column 2 shows the order of difficulty, from easiest to hardest, of the items within each Part of the examinations. Column 3 shows the order of difficulty of all the items in each examination. Thus the first question in Part I of the French examination is the fifth easiest of the 100 items in Part I, and the 10th easiest of the 275 questions in the whole French test. The fourth column shows for each item in each test the difference between the per cent of fourth-year and the per cent of second-year students that answered it correctly. In the case of the Physics test the difference is between the per cent of highest-quarter students and the per cent of lowest-quarter students that answered correctly. A minus sign before a figure in Column 4 indicates that a larger proportion of lower students answered the item correctly than of more advanced or better students.

The fifth column shows the order of "goodness" or of validity of all the items in each test.

The sixth and the seventh columns, appearing only in Part I of the French test, show the Henmon and the Wood frequencies of each French word in the test. The Henmon frequency means, e.g., that the first French word in Part I, *nom*, occurred 159 times in 400,000 words of running discourse; the Wood frequency means that the word *nom* was common to sixteen widely used French textbooks.

These columns of figures will be explained later in this report; for the moment the reader may confine his attention to the form, content and structure of the examinations. For convenience of reference the pages for each Part of each test are as follows:

French, Part I 210–216; Part II 217–221; Part III 222–229.
Spanish, Part I 230–235; Part II 236–240; Part III 241–249.
German, Part I 250–255; Part II 256–259; Part III 260–269.
Physics, (144 true-false statements) 270–282.

The reader will find it convenient to consult these references when the discussion of the columns of figures at the left of the questions is resumed below on page 296, and particularly on page 302.

[1] Cf. Junior High School Survey, above, pp. 9 to 38.

NOTE. *The reader must recall that, with the exception of the columns of italic figures at the left, these tests are reproduced here as they were presented to the students in June 1925, even to the inclusion of typographical errors, in addition to the linguistic infelicities that escaped our collaborators and proofreaders in that hurried spring of 1925. This is done so that the reader may have a first-hand view of one of the important factors in the experiment, and apprehend from concrete evidence the need for meticulous care in making examinations. These tests have been carefully corrected and revised with the coöperation of several modern language specialists and published as the Columbia Research Bureau Tests by the World Book Company, Yonkers, New York.*

REGENTS EXAMINATION IN FRENCH FOR JUNE 1925

SECOND HALF

2nd, 3rd and 4th Year Classes

Per cent correct answers	*Order of difficulty within parts*	*Order of difficulty whole test*	*Index of goodness*	*Order of goodness whole test*	*Henmon frequency*	*Wood frequency*
1	2	3	4	5	6	7
98.1	*5*	*10*	*3*	*253*	*159*	*16*
100.0	*1*	*1*	*0*	*267*	*1818*	*16*
97.1	*11*	*21*	*4*	*248*	*15*	*12*
95.4	*16*	*37*	*10*	*222*	*191*	*16*
98.3	*4*	*8*	*−2*	*269*	*39*	*15*

PART I — VOCABULARY

DIRECTIONS: *In each line select the English word or phrase which most nearly corresponds in meaning to the French word at the left, and put its number in the parentheses at the right, as in the examples. Do all the easy ones first. Twenty-five minutes.*

EXAMPLES:

menton 1 Chinese, 2 chin, 3 dog, 4 shine, 5 give.....(2)
dormir 1 door, 2 sleep, 3 give, 4 do, 5 go away.....(2)
quand 1 quarrel, 2 depart, 3 gone, 4 lost, 5 when.....(5)

1. nom 1 never, 2 name, 3 none, 4 write, 5 one.....( )
2. faire 1 makeshift, 2 affair, 3 make, 4 beautiful, 5 fairly...( )
3. feuille 1 leaf, 2 early, 3 furnish, 4 fearless, 5 fluid.....( )
4. comment 1 commence, 2 notice, 3 how, 4 commission, 5 many.( )
5. dimanche 1 sleeve, 2 eatable, 3 dismay, 4 Sunday, 5 image....( )

1	2	3	4	5	6	7			
98.4	*3*	*7*	*1*	*263*	*368*	*15*	6. cela	1 cell, 2 hers, 3 cease, 4 behold, 5 that	( )
92.0	*27*	*63*	*14*	*207*	*123*	*15*	7. dernier	1 last, 2 diurnal, 3 behind, 4 first, 5 endure	( )
91.0	*30*	*70*	*25*	*157*	*101*	*14*	8. prier	1 prime, 2 prior, 3 previous, 4 sacred, 5 pray	( )
86.6	*33*	*90*	*27*	*148*	*8*	*11*	9. fromage	1 rumor, 2 pity, 3 cheese, 4 yellow, 5 cherry	( )
94.5	*20*	*44*	*20*	*179*	*107*	*15*	10. presque	1 present, 2 almost, 3 press, 4 never, 5 previous	( )
95.3	*17*	*38*	*12*	*213*	*25*	*14*	11. réussir	1 useful, 2 succeed, 3 strengthen, 4 restrain, 5 successful	( )
92.4	*26*	*58*	*21*	*174*	*138*	*13*	12. apercevoir	1 receive, 2 pursuit, 3 notice, 4 billboard, 5 belong	( )
93.8	*21*	*50*	*16*	*197*	*229*	*15*	13. ainsi	1 into, 2 yearly, 3 always, 4 thus, 5 nearly	( )
46.7	*83*	*224*	*68*	*10*	*72*	*14*	14. cas	1 broken, 2 casket, 3 helmet, 4 site, 5 instance	( )
97.9	*6*	*12*	*2*	*257*	*69*	*16*	15. apporter	1 porter, 2 bring, 3 portable, 4 apportion, 5 entrance	( )
97.9	*7*	*13*	*2*	*258*	*105*	*14*	16. gros	1 rosy, 2 grovel, 3 small, 4 bad, 5 large	( )
95.6	*15*	*35*	*10*	*223*	*60*	*13*	17. bouche	1 full, 2 mouth, 3 pocket, 4 kiss, 5 hole	( )
98.5	*2*	*6*	*1*	*264*	*100*	*14*	18. chemin	1 churl, 2 cherish, 3 choose, 4 run, 5 route	( )
93.0	*25*	*56*	*8*	*229*	..	*11*	19. beurre	1 bread, 2 error, 3 burden, 4 burn, 5 butter	( )
97.2	*9*	*18*	*7*	*234*	*281*	*16*	20. laisser	1 break, 2 last, 3 leave, 4 bring, 5 least	( )
96.8	*12*	*25*	*3*	*254*	*111*	*14*	21. bientôt	1 later, 2 soon, 3 often, 4 well done, 5 bridal	( )
97.4	*8*	*17*	*4*	*249*	*36*	*11*	22. jambe	1 leg, 2 crowd, 3 yellow, 4 jam, 5 jump	( )
96.4	*13*	*28*	*7*	*235*	*110*	*13*	23. milieu	1 mirage, 2 miller, 3 place, 4 meadow, 5 middle	( )
96.2	*14*	*31*	*9*	*225*	*152*	*11*	24. côté	1 other, 2 cost, 3 side, 4 coat, 5 behind	( )
79.0	*46*	*125*	*45*	*61*	*76*	*15*	25. soin	1 care, 2 need, 3 wound, 4 suffering, 5 corner	( )

1	2	3	4	5	6	7		
97.1	*10*	*20*	*2*	*259*	*151*	*15*	26. droite	1 adroit, 2 drowsy, 3 dropped, 4 right, 5 placed.....( )
93.3	*24*	*54*	*12*	*214*	*8*	*4*	27. demoiselle	1 young lady, 2 deep, 3 guardian, 4 hostess, 5 demagogue.....( )
94.7	*19*	*42*	*14*	*208*	*11*	*5*	28. respirer	1 to restore, 2 withdraw, 3 to remove, 4 to long for, 5 to breathe.....( )
95.0	*18*	*41*	*6*	*239*	*7*	*4*	29. client	1 customer, 2 clinging, 3 pendulum, 4 clearing, 5 course.....( )
93.4	*23*	*53*	*13*	*210*	..	*4*	30. trompette	1 tramp, 2 horn, 3 small trunk, 4 footstool, 5 strength( )
57.4	*68*	*193*	*64*	*13*	*21*	*8*	31. cocher	1 driver, 2 to hide, 3 to coach, 4 to crow, 5 couch...( )
93.5	*22*	*52*	*6*	*240*	*12*	*8*	32. remplacer	1 to replace, 2 to reimburse, 3 to reveal, 4 to resemble, 5 success.....( )
80.0	*43*	*118*	*28*	*145*	..	*8*	33. blé	1 blessed, 2 in behalf of, 3 where, 4 wheat, 5 in spite of( )
91.4	*29*	*68*	*20*	*180*	..	*7*	34. repasser	1 to reveal, 2 to spy, 3 to go over again, 4 to retreat, 5 to criticise.....( )
88.4	*32*	*82*	*22*	*168*	*10*	*6*	35. marin	1 martial, 2 seaman, 3 marital, 4 mars, 5 moron....( )
83.3	*37*	*105*	*37*	*101*	*30*	*6*	36. obtenir	1 to hinder, 2 to threaten, 3 to procure, 4 obstacle 5 to hold against.....( )
89.8	*31*	*74*	*37*	*102*	..	*8*	37. crème	1 crime, 2 cream, 3 cheese, 4 wave, 5 crest.....( )
82.1	*38*	*110*	*40*	*85*	*44*	*8*	38. accomplir	1 complicate, 2 acclaim, 3 carry out, 4 accumulate, 5 make plain .....( )
67.7	*54*	*162*	*50*	*45*	*67*	*6*	39. son	1 bell, 2 deep, 3 sound, 4 soon, 5 boy.....( )
62.6	*61*	*180*	*61*	*23*	*9*	*7*	40. pis	1 foot, 2 pistol, 3 worse, 4 pedestrian, 5 peak.....( )

1	2	3	4	5	6	7		
83.7	*36*	*103*	*42*	*76*	*5*	*5*	41. amasser	1 to pack down, 2 to anoint, 3 to heap up, 4 to mash, 5 to destroy.....( )
91.6	*28*	*67*	*19*	*184*	*12*	*4*	42. venger	1 vinegar, 2 to avenge, 3 to conquer, 4 to arrive, 5 to convince.....( )
66.2	*55*	*165*	*55*	*30*	*10*	*4*	43. longueur	1 length, 2 languor, 3 grief, 4 longing, 5 lounger....( )
79.8	*45*	*120*	*32*	*123*	*8*	*4*	44. modifier	1 to modernize, 2 to alter, 3 to put in order, 4 painter, 5 clerk.....( )
74.0	*50*	*143*	*64*	*14*	*20*	*8*	45. éviter	1 to evict, 2 to avoid, 3 to invite, 4 to entertain, 5 to vitiate.....( )
80.3	*42*	*117*	*37*	*103*	..	*6*	46. poirier	1 fowl, 2 powder, 3 pear tree, 4 paper, 5 to purchase( )
86.5	*34*	*91*	*29*	*135*	..	*4*	47. noblesse	1 wound, 2 nausea, 3 nostalgia, 4 sickness, 5 greatness( )
65.4	*57*	*170*	*50*	*46*	*14*	*5*	48. satisfait	1 contented, 2 satisfaction, 3 satiated, 4 established, 5 support.....( )
80.4	*41*	*116*	*29*	*136*	*7*	*6*	49. loger	1 to box, 2 to lounge, 3 logging, 4 logical, 5 to lodge.( )
85.3	*35*	*96*	*17*	*193*	*20*	*4*	50. auberge	1 town, 2 inn, 3 twilight, 4 ambiguous, 5 large bed..( )
80.9	*40*	*114*	*16*	*198*	..	*5*	51. autrichien	1 clown, 2 Austrian, 3 magician, 4 bull-dog, 5 mechanical.....( )
60.7	*64*	*186*	*51*	*39*	*16*	*4*	52. confondre	1 to conform, 2 to conflict, 3 to compress, 4 to confuse, 5 to found.....( )
68.8	*53*	*160*	*40*	*86*	*16*	*7*	53. quartier	1 quart, 2 district, 3 hour, 4 money, 5 courtier.....( )
64.1	*58*	*176*	*60*	*25*	*19*	*8*	54. volontiers	1 volunteer, 2 willingly, 3 a kind of ship, 4 soldiers, 5 voluntary.....( )
59.1	*65*	*189*	*64*	*15*	*20*	*4*	55. piquer	1 piquant, 2 to pick up, 3 to sting, 4 to fish, 5 peach tree.....( )

1	2	3	4	5	6	7		
79.9	*44*	*119*	*47*	*51*	*21*	*8*	56. emploi	1 empty, 2 implore, 3 use, 4 train, 5 station.....( )
62.6	*62*	*181*	*34*	*115*	*21*	*4*	57. subitement	1 finally, 2 rapidly, 3 suddenly, 4 inevitable, 5 subjectively.....( )
62.9	*60*	*179*	*37*	*104*	*9*	*4*	58. rarement	1 scarcely, 2 rarefied, 3 rearrangement, 4 seldom, 5 gently.....( )
74.3	*49*	*142*	*62*	*22*	*84*	*8*	59. malgré	1 illness, 2 meager, 3 despite, 4 deeper, 5 despair...( )
75.5	*47*	*137*	*46*	*55*	*25*	*6*	60. ramasser	1 to leap up, 2 to fortify, 3 to collect, 4 ramify, 5 to pillage.....( )
71.5	*51*	*149*	*42*	*77*	..	*4*	61. démolir	1 to mollify, 2 to deprive, 3 to destroy, 4 to demote, 5 to modify.....( )
69.3	*52*	*158*	*44*	*67*	..	*5*	62. enricher	1 to enroll, 2 to make sad, 3 to run, 4 to embellish, 5 to alarm.....( )
58.8	*66*	*190*	*39*	*92*	*12*	*7*	63. paquet	1 boat, 2 wooden shoe, 3 packed, 4 package, 5 pasture.....( )
56.1	*69*	*197*	*63*	*18*	*21*	*4*	64. épée	1 thorn, 2 epic, 3 stick, 4 sword, 5 epidemic.....( )
62.4	*63*	*182*	*50*	*47*	*15*	*6*	65. peser	1 to fish, 2 to weigh, 3 to want, 4 poverty, 5 pouch..( )
55.6	*71*	*199*	*58*	*26*	*12*	*5*	66. légèrement	1 legibly, 2 less, 3 lightly, 4 leisurely, 5 legally.....( )
53.5	*75*	*207*	*51*	*40*	*35*	*8*	67. souffler	1 to suffocate, 2 to bellow, 3 to suffer, 4 to blow, 5 to slap.....( )
56.0	*70*	*198*	*63*	*19*	*8*	*7*	68. arroser	1 to touch, 2 to sprinkle, 3 to arrive, 4 to arise, 5 to spy.....( )
53.6	*74*	*206*	*31*	*126*	*46*	*8*	69. cloche	1 clock, 2 tower, 3 cloister, 4 bell, 5 lighthouse.....( )
34.2	*93*	*251*	*31*	*127*	*6*	*4*	70. nuire	1 nourish, 2 to injure, 3 walnut, 4 cloud, 5 nut.....( )

1	2	3	4	5	6	7		
54.8	*73*	*201*	*44*	*68*	*13*	*4*	71. botte	1 boat, 2 wine glass, 3 boot, 4 sandal, 5 bottom.....( )
51.0	*78*	*213*	*51*	*41*	*9*	*6*	72. abri	1 opening, 2 shelter, 3 apricot, 4 after, 5 morning...( )
39.2	*89*	*240*	*38*	*98*	..	*5*	73. poignet	1 point, 2 sword, 3 wrist, 4 fist, 5 poignant.....( )
40.8	*88*	*238*	*39*	*93*	*12*	*4*	74. égard	1 eager, 2 affection, 3 regard, 4 equal, 5 egregious...( )
55.0	*72*	*200*	*73*	*3*	*25*	*5*	75. arrière	1 carriage, 2 entrance court, 3 hallway, 4 back, 5 area ( )
51.2	*77*	*212*	*51*	*42*	*10*	*6*	76. écraser	1 to crush, 2 to erase, 3 to break, 4 to listen, 5 to cross( )
47.0	*82*	*221*	*55*	*31*	..	*5*	77. dépouiller	1 to strip, 2 to pull, 3 to deposit, 4 to deplore, 5 to disprove.....( )
24.3	*98*	*259*	*53*	*35*	..	*4*	78. disposé	1 dispossess, 2 ready, 3 unhappy, 4 disposition, 5 dispute.....( )
66.0	*56*	*167*	*−12*	*273*	*16*	*7*	79. sol	1 sun, 2 alone, 3 solitary, 4 soil, 5 solemn .....( )
49.2	*79*	*218*	*41*	*80*	..	*4*	80. douanier	1 charity-giver, 2 ship's captain, 3 customs officer, 4 foot-soldier, 5 donor.....( )
44.2	*84*	*228*	*57*	*28*	*5*	*4*	81. boue	1 buckle, 2 box, 3 mouth, 4 mud, 5 buoy.....( )
58.5	*67*	*192*	*70*	*6*	*22*	*4*	82. poussière	1 silk shop, 2 chicken yard, 3 basket, 4 dust, 5 bakery ( )
32.1	*94*	*251*	*16*	*199*	*10*	*5*	83. caisse	1 casing, 2 chest, 3 to cease, 4 caress, 5 he breaks..( )
48.1	*81*	*220*	*34*	*116*	..	*4*	84. allée	1 allied, 2 alley, 3 alleviated, 4 give, 5 allay.....( )
52.1	*76*	*209*	*54*	*33*	*9*	*5*	85. haleine	1 chain, 2 wool, 3 breath, 4 hamlet, 5 wild duck....( )
48.8	*80*	*219*	*44*	*69*	*19*	*4*	86. envahir	1 to evade, 2 to inveigh, 3 to avoid, 4 invidious, 5 to invade.....( )
26.6	*95*	*254*	*39*	*94*	*29*	*6*	87. éloigner	1 to elongate, 2 to elevate, 3 to listen, 4 distance, 5 to take away.....( )

1	2	3	4	5	6	7			
25.7	*96*	*256*	*17*	*194*	..	*6*	88.	chiffre	1 chiffonier, 2 rag, 3 cipher, 4 chip, 5 pepper.....( )
42.9	*86*	*231*	*47*	*52*	..	*4*	89.	pouls	1 chicken, 2 pool, 3 powder, 4 pull, 5 pulse.....( )
63.3	*59*	*178*	*15*	*202*	..	*4*	90.	farce	1 play, 2 distance, 3 funnel, 4 sheep, 5 fiction.....( )
37.9	*91*	*246*	*48*	*49*	*13*	*7*	91.	veiller	1 to be old, 2 to veil, 3 village, 4 to keep watch, 5 to vilify.....( )
43.9	*85*	*229*	*53*	*36*	*5*	*5*	92.	balai	1 bale, 2 brand, 3 balance, 4 skirt, 5 broom.....( )
36.7	*92*	*248*	*27*	*149*	*36*	*4*	93.	chasseur	1 driver, 2 foot-soldier, 3 shoe, 4 horse, 5 rabbit.....( )
16.4	*99*	*267*	*30*	*132*	*17*	*6*	94.	traîner	1 to train, 2 to board a train, 3 to trail, 4 to try, 5 to stretch.....( )
38.8	*90*	*243*	*27*	*150*	*12*	*8*	95.	coutume	1 cost, 2 count, 3 century, 4 costume, 5 custom.....( )
41.5	*87*	*236*	*70*	*7*	*16*	*6*	96.	mine	1 depth, 2 mean, 3 facet, 4 appearance, 5 mince....( )
25.2	*97*	*257*	*40*	*87*	*6*	*7*	97.	laboureur	1 worker, 2 elaborate, 3 plowman, 4 laborious, 5 labyrinth.....( )
10.7	*100*	*271*	*13*	*211*	..	*4*	98.	ardoise	1 ardent, 2 arduous, 3 difficult, 4 smooth, 5 slate....(
81.3	*39*	*112*	*−6*	*271*	..	*4*	99.	au-devant de	1 in front, 2 divan, 3 parasol, 4 audience, 5 toward..( )
74.4	*48*	*140*	*44*	*70*	*5*	*6*	100.	couverture	1 envelope, 2 green grass, 3 cover, 4 underestimate, 5 conversion.....( )

Per cent correct answers	*Order of difficulty within parts*	*Order of difficulty whole test*	*Index of goodness*	*Order of goodness whole test*
1	2	3	4	5
99.0	*1*	*2*	*0*	*268*
98.0	*6*	*11*	*2*	*260*
99.0	*2*	*3*	*5*	*243*
98.4	*5*	*9*	*5*	*244*
82.4	*51*	*109*	*37*	*105*
96.4	*13*	*29*	*5*	*245*
95.0	*18*	*40*	*7*	*236*
98.7	*3*	*4*	*3*	*255*
97.1	*10*	*22*	*9*	*226*
98.7	*4*	*5*	*1*	*265*
97.8	*7*	*14*	*2*	*261*
96.5	*11*	*26*	*11*	*220*
97.1	*9*	*19*	*9*	*227*
93.0	*29*	*64*	*12*	*215*
94.5	*19*	*45*	*12*	*216*

PART II — COMPREHENSION

DIRECTIONS: *The following statements in French are either true or false. Indicate by a plus sign (+) a true statement, and by a zero (0) a false statement. Put the plus signs and zeros in the parentheses at the right. Do not spend too much time on any one statement. Do the easy ones first. Twenty minutes.*

1. Le dimanche est un jour de travail aux États-Unis mais le lundi on se repose. ( )
2. Ordinairement on travaille le jour et on se repose la nuit. ( )
3. Les pères et les mères des soldats et des généraux sont heureux pendant une bataille ( )
4. Les parents sont joyeux et fiers du succès de leurs enfants. ( )
5. L'homme riche dont la bourse est fermée aux malheureux est un être qu'on n'admire pas souvent. ( )
6. Au Canada, il neige souvent en hiver. ( )
7. La vie de l'homme est plus longue que celle du chien ( )
8. La sœur de mon père est ma tante. ( )
9. Tout le monde dit que la guerre est bonne pour la santé ( )
10. Une minute est la soixantième partie d'une heure ( )
11. Il ne faut pas toujours dire ce que l'on pense mais il est bon de penser avant de parler. ( )
12. Le chapeau se porte sur les épaules. ( )
13. On sert le déjeuner tout de suite après le dîner. ( )
14. Quatre heures de sommeil chaque jour est assez pour les enfants ( )
15. On trouve dans l'histoire ancienne une foule de faits intéressants. ( )

1	2	3	4	5	
95.9	*17*	*34*	*4*	*250*	16. Les forêts et les champs sont verts au printemps.....( )
94.3	*20*	*46*	*18*	*189*	17. Quand quelqu'un frappe à la porte il faut toujours fermer la porte à clef sans demander qui est là.....( )
93.0	*30*	*65*	*25*	*158*	18. De petites fortunes réunies en font une grande.....( )
92.4	*26*	*59*	*27*	*151*	19. Les vrais amis nous restent fidèles dans le malheur aussi bien que dans le bonheur.....( )
67.3	*64*	*164*	*31*	*128*	20. L'ours est un animal qui ne vit que dans les pays chauds.....( )
96.5	*12*	*27*	*5*	*246*	21. Il y a beaucoup d'enfants qui n'aiment pas aller à l'école.....( )
93.7	*23*	*51*	*20*	*181*	22. Les fruitiers ne vendent que des pommes.....( )
88.9	*38*	*79*	*20*	*182*	23. Tous les quatre ans seulement il y a un vingt-neuf février.....( )
92.1	*28*	*62*	*11*	*221*	24. Ceux qui partent les premiers n'arrivent pas toujours les premiers.....( )
90.9	*31*	*71*	*21*	*175*	25. Les chaises sont des objets presque inutiles.....( )
89.3	*35*	*76*	*42*	*78*	26. Beaucoup de sous ensemble font souvent une somme importante.....( )
92.4	*27*	*60*	*23*	*163*	27. Toutes les maisons américaines n'ont que cinq chambres.....( )
97.8	*8*	*15*	*4*	*251*	28. On plante toutes les fleurs en hiver.....( )
88.3	*40*	*83*	*23*	*164*	29. Les poissons vivent dans l'eau.....( )
93.9	*21*	*48*	*17*	*195*	30. On va plus vite en marchant qu'en courant.....( )
89.0	*37*	*78*	*27*	*152*	31. Une république est fière de son roi.....( )
89.4	*34*	*75*	*30*	*133*	32. On ne peut pas être en plusieurs endroits à la fois.....( )
18.8	*75*	*263*	*−17*	*274*	33. Une personne peut avoir l'air dur tout en étant très bonne.....( )
87.0	*41*	*86*	*6*	*241*	34. Riches ou pauvres, nous sommes tous égaux devant la mort.....( )
96.3	*14*	*30*	*−2*	*270*	35. Les Françaises ne parlent jamais le français à la campagne.....( )

1	2	3	4	5		
83.7	*49*	*102*	*19*	*185*	36. Le soleil se lève à l'ouest	( )
77.7	*58*	*128*	*29*	*137*	37. Peu d'enfants se rendent compte du travail qui est nécessaire pour produire un morceau de pain	( )
71.3	*62*	*150*	*41*	*81*	38. Il y a des pendules qui sonnent douze heures à minuit	( )
69.0	*63*	*159*	*27*	*153*	39. La guerre n'était pas faite autrefois comme de nos jours	( )
75.4	*61*	*138*	*21*	*176*	40. Tous ceux qui habitent la campagne sont des paysans	( )
77.6	*59*	*129*	*25*	*159*	41. Celui qui fait le pain s'appelle un boulanger	( )
78.7	*57*	*126*	*26*	*154*	42. La poire est un légume délicieux	( )
90.2	*33*	*73*	*7*	*237*	43. Quand on veut faire un voyage on choisit toujours le mois de juillet	( )
96.1	*15*	*32*	*8*	*230*	44. S'il pleut, il faut aller se promener	( )
93.9	*22*	*49*	*18*	*190*	45. Il n'y a que deux mois de trente jours chacun	( )
85.1	*45*	*97*	*30*	*134*	46. On voit les étoiles seulement quand il y a pleine lune	( )
76.0	*60*	*135*	*38*	*99*	47. Un garçon met son chapeau dans la maison mais il l'ôte chaque fois qu'il sort	( )
63.7	*68*	*177*	*46*	*56*	48. On lave le linge parce qu'il est propre	( )
66.0	*66*	*168*	*46*	*57*	49. L'or et l'argent sont des métaux peu chers	( )
66.2	*65*	*166*	*67*	*11*	50. 'Minuit' veut dire soixante secondes	( )
82.0	*52*	*111*	*44*	*71*	51. Ce n'est qu'en hiver qu'on voit beaucoup d'oiseaux dans les arbres	( )
84.0	*48*	*101*	*10*	*224*	52. Les mauvais arbres ne portent point de bons fruits	( )
90.9	*32*	*72*	*23*	*165*	53. Les hommes ont toujours cru que la terre est ronde	( )
79.6	*55*	*121*	*56*	*29*	54. Les enfants ne peuvent se blesser qu'au genou en tombant	( )
87.0	*42*	*87*	*12*	*217*	55. Les monuments ne servent à rien	( )

1	2	3	4	5	
85.6	*43*	*94*	*28*	*146*	56. Il est tout à fait injuste que l'on punisse les bons.....( )
92.9	*25*	*57*	*15*	*203*	57. L'Amérique fut découverte par Christophe Colomb en quatorze cent quatre-vingt douze.....( )
85.1	*46*	*98*	*26*	*155*	58. Les plus vieilles amitiés sont généralement les plus douces.....( )
93.9	*39*	*81*	*29*	*138*	59. Chaque siècle se compose de cinquante ans.....( )
83.2	*50*	*106*	*31*	*129*	60. L'après-midi est la partie du jour entre le coucher du soleil et minuit.....( )
89.3	*36*	*77*	*12*	*218*	61. Il y a plusieurs fleuves dans notre pays.....( )
80.9	*53*	*113*	*34*	*117*	62. Pour couper la viande, il ne faut que des couteaux qui ne coupent pas.....( )
85.5	*44*	*95*	*48*	*50*	63. Le devoir de vos parents est de vous obéir et le vôtre est de les guider.....( )
61.1	*70*	*185*	*29*	*139*	64. Celui qui convoite les biens d'autrui, cherche quelquefois à les lui prendre injustement.....( )
93.3	*24*	*55*	*16*	*200*	65. Les chevaux ne peuvent pas travailler toujours sans se reposer.....( )
96.1	*16*	*33*	*7*	*238*	66. Quand on n'a pas réfléchi avant de parler on regrette quelquefois ce que l'on a dit.....( )
85.2	*47*	*99*	*22*	*169*	67. En temps de paix on n'a pas besoin de servir son pays.....( )
79.2	*56*	*124*	*22*	*170*	68. Il est utile de dater les lettres que nous écrivons.....( )
65.3	*67*	*171*	*40*	*88*	69. Celui qui reconnaît ses défauts n'est pas nécessairement sur la voie du salut ( )
80.9	*54*	*115*	*16*	*201*	70. Il y a des hommes qui doivent la considération dont ils jouissent à leur fortune et non à leur mérite.....( )
61.8	*69*	*183*	*43*	*75*	71. Les voleurs qui, de nos jours, arrivent à dérober des tableaux de prix ne sont pas à craindre dans les musées.....( )

1	2	3	4	5	
53.8	*71*	*205*	*−37*	*275*	72. Quoi que l'on dise et quoi que l'on pense, c'est bien sur l'apparence qu'il faut juger le caractère .....( )
42.2	*73*	*240*	*54*	*34*	73. On peut se permettre de jurer qu'on ne fera jamais ceci ou cela, car on n'est pas sans savoir ce que l'avenir lui réserve.....( )
43.3	*72*	*230*	*40*	*89*	74. Gaspiller le pain est très louable, même lorsqu'on est entouré de malheureux qui en ont grand besoin.....( )
31.6	*74*	*253*	*33*	*120*	75 Le fat croit que le luxe seul est capable d'attirer l'attention sur lui; mais il n'y a que le savant qui puisse être tenté de l'admirer .....( )

PART III — GRAMMAR

DIRECTIONS: *Each of the English sentences below is followed by a translation which is correct as far as it goes, but which is incomplete. Each translation can be completed correctly by inserting one, two, or three words in each blank* (——). *When you have decided what is necessary to complete the translation, write it on the line at the right of the translation. WRITE LEGIBLY. Do not spend too much time on any one sentence. Do all the easy ones first, then go back to the harder ones if you have time.*

Write carefully and clearly, putting your answers on the lines at the right, as in the examples. Forty-five minutes.

EXAMPLES:

1. I see some dogs. Je vois (——) chiens ..... ——des
2. I am hungry. (——) faim ..... ——J'ai
3. I make him read this book. Je lui (——) ce livre..... ——fais lire

Per cent correct answers	*Order of difficulty within parts*	*Order of difficulty whole test*	*Index of goodness*	*Order of goodness whole test*	
1	2	3	4	5	
95.6	*4*	*36*	*8*	*231*	1. This man is my brother. (——) homme est mon frère.. ———
96.9	*2*	*23*	*2*	*262*	2. He is taller than John. Il est plus grand (——) Jean.. ———
96.9	*3*	*24*	*12*	*219*	3. Give me your hand. (——) la main..... ———
94.6	*6*	*43*	*8*	*232*	4. I have finished my lessons. J'ai (——) mes leçons.... ———
94.1	*7*	*47*	*13*	*212*	5. Who are you? (——) êtes-vous?..... ———
95.2	*5*	*39*	*6*	*242*	6. We have struck them. Nous (——) avons frappés.... ———
86.9	*15*	*88*	*29*	*140*	7. What pen have you? (——) plume avez-vous?..... ———
97.7	*1*	*16*	*3*	*256*	8. His sister and he are there. Sa sœur et lui (——) là... ———

1	2	3	4	5		
91.4	*10*	*68*	*19*	*186*	9. I speak to the men. Je parle (——) nommes.....	.....
91.7	*9*	*65*	*9*	*228*	10. I speak of John. Je parle (——) Jean.....	.....
86.2	*17*	*92*	*4*	*252*	11. We drink hot coffee. Nous buvons du café (——).....	.....
84.9	*18*	*99*	*35*	*111*	12. I see my friend. Je vois (——) amie.....	.....
76.1	*29*	*133*	*38*	*100*	13. Tell me her story. Racontez-moi (——) histoire.....	.....
87.3	*13*	*84*	*20*	*183*	14. Does that man speak English? Cet homme (——) anglais?.....	.....
87.9	*12*	*83*	*18*	*191*	15. Give it to her. Donnez (——).....	.....
92.3	*8*	*60*	*15*	*204*	16. Give him the pen. Donnez (——) la plume.....	.....
77.6	*24*	*126*	*33*	*121*	17. At what o'clock will he come? (——) heure viendra-t-il?	.....
75.2	*31*	*145*	*29*	*141*	18. He has those pens. Il a (——) plumes-là.....	.....
86.3	*16*	*91*	*22*	*130*	19. See the great kings. Voyez les (——) rois.....	.....
70.7	*40*	*152*	*46*	*58*	20. He has the pupil's pen. Il a la plume (——) élève....	.....
87.0	*14*	*87*	*36*	*109*	21. Life is short. (——) est courte.....	.....

1	2	3	4	5		
83.3	*20*	*106*	*31*	*130*	22. She has her pencil and mine. Elle a son crayon et (——)	
79.6	*22*	*121*	*21*	*177*	23. He has enough meat. Il a assez (——) viande	
83.7	*19*	*103*	*21*	*178*	24. I give her the thimble with pleasure. Je (——) donne le dé avec plaisir	
77.0	*26*	*130*	*31*	*131*	25. Give me my book and pen. Donnez-moi mon livre (——) plume	
88.9	*11*	*79*	*37*	*106*	26. He needs a pen. Il a besoin (——) une plume	
68.3	*45*	*160*	*29*	*142*	27. The wall is high. Le mur est (——)	
64.5	*51*	*175*	*23*	*166*	28. The general was closely followed by the enemy. Le général fut suivi de près (——) l'ennemi	
71.3	*38*	*150*	*51*	*43*	29. Do I speak of him? (——) que je parle de lui?	
70.1	*43*	*155*	*25*	*160*	30. He is favorable to my projects. Il est favorable (——) projets	
73.7	*35*	*145*	*33*	*122*	31. It has often happened. Cela (——) souvent arrivé	
82.9	*21*	*107*	*32*	*124*	32. If he has come, let me know. S'il (——) venu, faites-le-moi savoir	
73.0	*36*	*146*	*37*	*107*	33. I like apples and pears. J'aime les pommes et (——)	
60.2	*54*	*188*	*22*	*172*	34. Pass your hand over this cloth. Passez la main (——) ce drap	

1	2	3	4	5		
61.4	*52*	*184*	*49*	*48*	35. You are very kind. Vous êtes (——).....	.....
79.6	*23*	*122*	*41*	*82*	36. He has neither father nor mother. Il n'a (——) père (——) mère.....	.....
71.2	*39*	*151*	*45*	*62*	37. He walks faster than I thought. Il marche (——) que je ne pensais.....	.....
51.3	*64*	*203*	*47*	*53*	38. The horse is the friend of man. Le cheval est l'ami (——)	.....
70.7	*41*	*153*	*44*	*72*	39. How long have you been waiting? Depuis quand (——) vous?.....	.....
76.8	*27*	*131*	*23*	*167*	40. I will not remain. Je ne veux pas (——).....	.....
60.2	*53*	*187*	*47*	*54*	41. I have some of your money. J'ai (——) argent.....	.....
46.9	*69*	*222*	*53*	*37*	42. Have you any money about you? Avez-vous de l'argent (——) vous?.....	.....
74.0	*33*	*143*	*45*	*63*	43. He gives you his hand. Il vous donne (——) main....	.....
69.4	*44*	*156*	*72*	*4*	44. You are wanted. (——) vous demande.....	.....
75.9	*30*	*135*	*35*	*112*	45. He speaks of going to Paris. Il parle (——) à Paris...	.....
25.1	*87*	*258*	*22*	*173*	46. Tomorrow will be the sixteenth. (——) sera demain le seize.....	.....
57.8	*55*	*192*	*82*	*1*	47. He has sore eyes. Il a mal (——) yeux.....	.....

1	2	3	4	5		
74.6	*32*	*140*	*32*	*125*	48. She arrived in the spring. Elle arriva (——) printemps	.....
53.0	*62*	*208*	*19*	*187*	49. Is that true? Cela (——) vrai?.....	.....
38.8	*80*	*264*	*29*	*143*	50. He is easy to satisfy. Il est facile (——) contenter....	.....
76.7	*28*	*132*	*35*	*113*	51. Mary and Louisa have come. Marie et Louise sont (——).....	.....
50.4	*66*	*215*	*45*	*64*	52. She is the least amiable. Elle est (——) moins aimable	.....
65.0	*49*	*173*	*66*	*12*	53. There is somebody at the door. Il y a (——) à la porte	.....
71.9	*37*	*147*	*28*	*147*	54. He is at Mr. Ribot's. Il (——) Monsieur Ribot.....	.....
51.4	*63*	*210*	*46*	*59*	55. I should have gone, had he come. Je (——) parti, s'il était venu.....	.....
65.1	*48*	*140*	*61*	*24*	56. Queen Victoria is beloved. (——) Victoria est aimée..	.....
77.6	*25*	*129*	*5*	*247*	57. However brave he thinks himself. Si brave qu'il se (——)	.....
74.0	*34*	*144*	*36*	*110*	58. You will have to stay. Il faudra que vous (——).....	.....
65.8	*47*	*169*	*39*	*95*	59. I shall lend them the books. Je (——) prêterai les livres	.....
50.7	*65*	*214*	*45*	*65*	60. Give me some of those pens. Donnez-moi (——) ces plumes-là.....	.....

1	2	3	4	5		
38.7	*81*	*244*	*70*	*8*	61. There are about 100 pupils. Il y a une centaine (——)	
67.7	*46*	*162*	*64*	*16*	62. However great you may be! Quelque grand que vous (——)!	
45.4	*71*	*225*	*71*	*5*	63. He was covered with wounds. Il était couvert (——) blessures	
54.5	*60*	*203*	*29*	*144*	64. Mary has some pretty flowers. Marie a (——) jolies fleurs	
42.1	*74*	*233*	*24*	*162*	65. I passed through the forest. J'ai passé (——) la forêt	
34.8	*85*	*250*	*15*	*205*	66. They have washed their hands. Ils se sont (——) les mains	
70.7	*42*	*154*	*25*	*161*	67. I give them to them. Je les (——) donne	
42.1	*75*	*234*	*64*	*17*	68. Most of the soldiers perished. La (——) soldats périrent	
64.6	*50*	*174*	*37*	*108*	69. To hesitate would be weakness. Hésiter (——) faiblesse	
22.8	*88*	*260*	*44*	*73*	70. I shall come presently. Je viendrai (——) l'heure	
36.0	*84*	*249*	*76*	*2*	71. Who tells such a story? Qui raconte (——) histoire?	
38.9	*79*	*241*	*58*	*27*	72. Many generals have come here. Beaucoup de (——) y sont venus	
38.5	*82*	*245*	*26*	*156*	73. Give me some of the same. Donnez-moi (——) mêmes	

1	2	3	4	5		
42.0	*69*	*235*	*42*	*79*	74. Whose son are you? (——) êtes-vous fils?	
46.8	*70*	*223*	*69*	*9*	75. Can you not help me? Ne sauriez vous (——)?	
44.4	*73*	*227*	*63*	*20*	76. Tell them to be ready. Dis-leur qu'ils (——) prêts	
56.3	*57*	*195*	*44*	*74*	77. I owe to him that I am still alive. Je lui dois (——) encore en vie	
49.8	*67*	*216*	*39*	*96*	78. I think it is raining. Je crois (——)	
56.1	*58*	*196*	*51*	*44*	79. If I had had time, I should have gone to London. Si j'avais eu le temps, je (——) allé à Londres	
18.7	*9*	*264*	*34*	*118*	80. I cannot tell you. Je ne (——) le dire	
37.1	*83*	*247*	*−7*	*272*	81. He is very clever in business. Il est bien habile (——) affaires	
49.7	*68*	*217*	*46*	*60*	82. It does not seem to me he is mad. Il ne me semble pas qu'il (——) fou	
44.6	*72*	*226*	*39*	*97*	83. The women wept, screamed and gesticulated. Les femmes pleuraient, (——) gesticulaient	
26.3	*86*	*255*	*17*	*197*	84. I know both the father and son. Je connais (——) le père (——) le fils	
14.9	*95*	*269*	*40*	*90*	85. They are pleased with you. Elles sont (——) de vous	
21.1	*90*	*262*	*52*	*38*	86. I shall have a letter written. Je (——) une lettre	

1	2	3	4	5	
56.5	*56*	*194*	*18*	*192*	87. Do you not fear he will come? Ne craignez-vous pas qu'il (——)? .....
13.7	*96*	*270*	*14*	*209*	88. That happened from carelessness. Cela arriva (——) négligence .....
41.0	*77*	*237*	*45*	*66*	89. He divided it among his friends. Il le partagea (——) amis
54.5	*61*	*204*	*34*	*119*	90. After he had dined, he set out. Après qu'il (——) dîné, il partit .....
39.3	*78*	*239*	*55*	*32*	91. They speak of one another. Ils parlent les uns (——) autres .....
16.3	*94*	*268*	*35*	*114*	92. Though I doubted that he would come. Quoique je doutasse qu'il (——) .....
22.0	*89*	*261*	*40*	*91*	93. I thought he was coming back. Je croyais qu'il (——)
54.7	*59*	*169*	*63*	*21*	94. You ought to speak to him. Vous (——) lui parler ....
16.6	*93*	*266*	*41*	*83*	95. The queen was followed by her ladies. La reine était (——) ses dames .....
5.5	*97*	*272*	*19*	*171*	96. He showed extreme care. Il montra (——) soin extrême
18.3	*92*	*265*	*41*	*84*	97. Seeing is believing. Voir (——) croire .....
4.6	*98*	*273*	*15*	*206*	98. When you have come home, I shall go out. Quand vous (——) je sortirai .....
2.5	*99*	*274*	*8*	*233*	99. He has patience equal to anything. Il (——) patience à toute épreuve .....
2.0	*100*	*275*	*1*	*266*	100. I was in a temper. J'étais (——) humeur .....

REGENTS EXAMINATION IN SPANISH FOR JUNE 1925

SECOND HALF

2nd, 3rd and 4th Year Classes

PART I — VOCABULARY

DIRECTIONS: *In each line select the English word or phrase which most nearly corresponds in meaning to the Spanish word at the left, and put its number in the parentheses at the right, as in the examples. Do all the easy ones first. Twenty-five minutes.*

EXAMPLES:

libro 1 paper, 2 book, 3 pencil, 4 free, 5 make.....(2)
papel 1 popular, 2 read, 3 fly, 4 pray, 5 paper.....(5)
escribir 1 scribe, 2 play, 3 write, 4 pen, 5 ink.....(3)

Per cent correct answers	*Order of difficulty within parts*	*Order of difficulty whole test*	*Index of goodness*	*Order of goodness whole test*			
1	2	3	4	5			
100.0	*1*	*1*	*4*	*248*	1. vivir	1 live, 2 energize, 3 tack, 4 visit, 5 revive	( )
98.7	*6*	*13*	*5*	*234*	2. andar	1 dive, 2 give out, 3 seek, 4 walk, 5 pathway	( )
99.0	*4*	*10*	*4*	*257*	3. mujer	1 woman, 2 major, 3 mule, 4 brown, 5 servant	( )
98.4	*10*	*17*	*3*	*262*	4. repetir	1 repent, 2 report, 3 replace, 4 compete, 5 repeat	( )
99.7	*2*	*4*	*4*	*249*	5. palabra	1 match, 2 parley, 3 hard work, 4 word, 5 straw	( )
98.4	*11*	*18*	*4*	*250*	6. reír	1 laugh, 2 reign, 3 right, 4 river, 5 rain	( )
99.7	*3*	*5*	*4*	*251*	7. enemigo	1 enemy, 2 indigo, 3 energize, 4 friendship, 5 ensign	( )
98.4	*12*	*19*	*4*	*252*	8. cabeza	1 cabbage, 2 gentleman, 3 horse, 4 head, 5 hair	( )
98.7	*8*	*14*	*5*	*235*	9. lápiz	1 stone, 2 blackboard, 3 lap, 4 notebook, 5 pencil	( )
97.0	*17*	*32*	*6*	*214*	10. cara	1 cart, 2 care, 3 face, 4 deer, 5 meat	( )
92.5	*26*	*65*	*1*	*272*	11. ya	1 still, 2 already, 3 once, 4 yes, 5 yellow	( )
96.4	*18*	*37*	*9*	*171*	12. criado	1 servant, 2 tearful, 3 shriek, 4 unripe, 5 call	( )

1	2	3	4	5			
88.9	*43*	*93*	*14*	*118*	13.	legumbre	1 tearful, 2 tongue, 3 beans, 4 big horn, 5 vegetable ..... ( )
98.4	*13*	*20*	*4*	*258*	14.	creer	1 create, 2 believe, 3 make, 4 cross, 5 crease ..... ( )
90.9	*36*	*81*	*11*	*147*	15.	hora	1 now, 2 today, 3 then, 4 hour, 5 house ..... ( )
95.1	*21*	*44*	*7*	*202*	16.	sur	1 south, 2 on, 3 sister, 4 sure, 5 surety ..... ( )
89.8	*39*	*88*	*14*	*119*	17.	bosque	1 moreover, 2 besides, 3 search, 4 push, 5 forest ..... ( )
91.8	*29*	*72*	*8*	*187*	18.	cuarto	1 square, 2 room, 3 four, 4 box, 5 floor ..... ( )
88.5	*44*	*95*	*10*	*153*	19.	torre	1 torrent, 2 tower, 3 bull, 4 twist, 5 narrow ..... ( )
95.1	*22*	*45*	*7*	*192*	20.	bastante	1 great aunt, 2 baseness, 3 insufficient, 4 rather, 5 enough .... ( )
93.1	*24*	*59*	*6*	*215*	21.	pedir	1 lose, 2 pay for, 3 weigh, 4 ask for, 5 go afoot ..... ( )
99.0	*5*	*11*	*5*	*236*	22.	terminar	1 journey, 2 end, 3 run into the ground, 4 pause, 5 term ..... ( )
95.3	*20*	*45*	*8*	*182*	23.	busca	1 bush, 2 burn, 3 sea-bird, 4 search, 5 seed ..... ( )
97.3	*15*	*30*	*7*	*193*	24.	gobierno	1 government, 2 politics, 3 Gothic, 4 wisdom, 5 justice ..... ( )
88.5	*45*	*96*	*7*	*203*	25.	sentir	1 scant, 2 seat, 3 regret, 4 sentiment, 5 help ..... ( )
92.1	*28*	*69*	*10*	*154*	26.	llorar	1 lower, 2 carry, 3 weep, 4 rain, 5 gain ..... ( )
98.7	*8*	*15*	*5*	*243*	27.	domingo	1 Sunday, 2 holiday, 3 Saturday, 4 domino, 5 Easter ..... ( )
90.9	*37*	*82*	*7*	*194*	28.	permiso	1 promise, 2 possession, 3 perhaps, 4 permission, 5 exchange .. ( )
91.2	*33*	*76*	*5*	*244*	29.	llevar	1 fill, 2 carry, 3 bring, 4 call, 5 lift ..... ( )
98.4	*14*	*21*	*6*	*216*	30.	olvidar	1 empty, 2 avoid, 3 invite, 4 forget, 5 lose out ..... ( )
90.2	*38*	*86*	*12*	*139*	31.	toro	1 tower, 2 fierce, 3 small, 4 bull, 5 torrent ..... ( )
93.4	*23*	*55*	*8*	*177*	32.	destruir	1 distribute, 2 hate, 3 wish for, 4 disdain, 5 destroy ..... ( )
61.4	*89*	*206*	*18*	*66*	33.	almohada	1 blanket, 2 haste, 3 breakfast, 4 pillow, 5 carpet ..... ( )

1	2	3	4	5			
72.5	*77*	*171*	*2*	*268*	34. devolver	1 resolve, 2 disturb, 3 return, 4 develop, 5 fly back.....( )
89.5	*41*	*90*	*6*	*224*	35. sobretodo	1 overcoat, 2 nephew, 3 sobriety, 4 superfluous, 5 singularity.( )
98.7	*9*	*16*	*6*	*217*	36. a la izquierda	1 to one side, 2 near by, 3 surrounding, 4 on the left, 5 hard by.( )
83.6	*57*	*124*	*27*	*14*	37. mostrar	1 devote, 2 sarcastic, 3 mount, 4 go up, 5 show.....( )
69.9	*79*	*178*	*17*	*85*	38. par	1 equality, 2 part, 3 pair, 4 place, 5 several.....( )
79.4	*65*	*148*	*21*	*40*	39. impedir	1 to be important, 2 prevent, 3 insult, 4 compete, 5 imitate..( )
86.9	*48*	*103*	*−12*	*275*	40. librero	1 liberator, 2 bookseller, 3 livery, 4 laborer, 5 foreman.....( )
91.5	*31*	*74*	*16*	*66*	41. azúcar	1 to cut, 2 irritate, 3 lunch, 4 sugar, 5 light.....( )
85.3	*52*	*111*	*20*	*49*	42. equivocarse	1 to be mistaken, 2 question, 3 disgust, 4 pollute, 5 cry out..( )
84.3	*54*	*116*	*19*	*54*	43. retirarse	1 to collect, 2 scold, 3 return, 4 withdraw, 5 distribute.....( )
91.8	*30*	*73*	*13*	*126*	44. confianza	1 confidence, 2 happiness, 3 conference, 4 influenza, 5 confection.....( )
86.2	*49*	*105*	*6*	*225*	45. manzana	1 pig, 2 tomorrow, 3 slaughter, 4 plum, 5 apple.....( )
96.1	*19*	*40*	*7*	*204*	46. alegrarse	1 light, 2 lift, 3 call to, 4 happiness, 5 be glad.....( )
91.5	*32*	*75*	*13*	*127*	47. panadería	1 panic, 2 paper-mill, 3 bakery, 4 cannery, 5 meat shop.....( )
87.9	*46*	*100*	*12*	*140*	48. espera	1 thickness, 2 expectation, 3 agitation, 4 appearance, 5 hate..( )
91.1	*34*	*77*	*9*	*165*	49. compuesto	1 pressed, 2 ended, 3 composed, 4 placed upon, 5 compelled.( )
81.0	*63*	*140*	*10*	*155*	50. jamás	1 still, 2 either, 3 no one, 4 never, 5 not any.....( )
91.1	*35*	*78*	*10*	*158*	51. estrella	1 stairway, 2 stretch, 3 strict, 4 trellis, 5 star.....( )
89.2	*42*	*92*	*16*	*105*	52. cesar	1 stop, 2 marry, 3 kiss, 4 Caesar, 5 fall.....( )
87.2	*47*	*102*	*18*	*67*	53. además	1 slowly, 2 often, 3 moreover, 4 rather, 5 after.....( )

1	2	3	4	5			
67.3	*81*	*187*	*26*	*18*	54. recordar	1 cover up, 2 receive, 3 record, 4 recover, 5 remember	( )
84.0	*56*	*122*	*17*	*86*	55. sombra	1 hat, 2 sobriety, 3 sadness, 4 shadow, 5 dark	( )
84.9	*53*	*112*	*7*	*205*	56. ladrón	1 heavy load, 2 dog, 3 slave, 4 laziness, 5 robber	( )
97.4	*16*	*31*	*9*	*166*	57. corazón	1 heart, 2 courage, 3 correction, 4 reason, 5 hearty	( )
62.4	*87*	*203*	*14*	*120*	58. junto a	1 joined together, 2 in order that, 3 hunt, 4 point, 5 next to	( )
72.9	*75*	*187*	*24*	*23*	59. ensalada	1 insulated, 2 greens, 3 salad, 4 assortment, 5 exorbitant	( )
76.4	*68*	*157*	*20*	*54*	60. deber	1 doubt, 2 gift, 3 duty, 4 finger, 5 weakness	( )
83.7	*58*	*125*	*16*	*99*	61. mitad	1 mitten, 2 error, 3 might, 4 half, 5 method	( )
82.0	*62*	*136*	*14*	*115*	62. primo	1 prim, 2 principal, 3 prime, 4 first, 5 cousin	( )
71.2	*78*	*175*	*5*	*237*	63. gallo	1 expense, 2 eat, 3 rooster, 4 chicken, 5 gallop	( )
82.3	*61*	*134*	*19*	*55*	64. tostada	1 bread, 2 toast, 3 loaf, 4 rudeness, 5 work	( )
84.3	*55*	*117*	*12*	*141*	65. oveja	1 eye, 2 egg, 3 sheep, 4 old, 5 distant	( )
79.7	*64*	*142*	*16*	*100*	66. luego	1 carry, 2 mournful, 3 place, 4 now, 5 then	( )
83.0	*60*	*130*	*18*	*68*	67. cocer	1 catch, 2 concede, 3 cook, 4 drive, 5 coax	( )
86.2	*50*	*106*	*−4*	*274*	68. nube	1 nob, 2 cloud, 3 knot, 4 now, 5 no one	( )
57.5	*97*	*235*	*24*	*25*	69. acá	1 near, 2 yonder, 3 in that, 4 hither, 5 there	( )
63.8	*86*	*198*	*13*	*128*	70. lejos	1 distance, 2 far, 3 laws, 4 bed, 5 lies	( )
92.4	*27*	*66*	*16*	*93*	71. viuda	1 neighbor, 2 seen, 3 widow, 4 wine, 5 pension	( )
76.1	*69*	*159*	*28*	*11*	72. convencer	1 to be convenient, 2 convince, 3 conviction, 4 conform, 5 enforce	( )
66.7	*82*	*190*	*8*	*188*	73. cantidad	1 song, 2 quantity, 3 carefully, 4 quality, 5 candidate	( )

1	2	3	4	5			
44.4	*95*	*232*	*26*	*19*	74. mandar	1 bind, 2 send, 3 demand, 4 manufacture, 5 give up	( )
69.6	*80*	*180*	*27*	*15*	75. cansar	1 canopy, 2 fatigue, 3 forgive, 4 endure, 5 cancer	( )
72.9	*76*	*169*	*19*	*79*	76. mejorar	1 stir, 2 measure, 3 deserve, 4 improve, 5 prescribe	( )
83.7	*59*	*126*	*26*	*16*	77. sobrino	1 hat, 2 dark, 3 nephew, 4 cousin, 5 heavy	( )
48.4	*93*	*227*	*26*	*17*	78. sencillo	1 armchair, 2 soft, 3 eyebrow, 4 simple, 5 sensible	( )
51.3	*91*	*223*	*22*	*35*	79. nombrar	1 number, 2 crowd, 3 mention, 4 push, 5 refund	( )
77.4	*67*	*154*	*20*	*50*	80. vacío	1 empty, 2 vacation, 3 vaccination, 4 cow, 5 value	( )
65.7	*84*	*193*	*12*	*142*	81. barrer	1 sweep, 2 bear, 3 erase, 4 bearer, 5 bathe	( )
89.9	*40*	*89*	*11*	*144*	82. juego	1 young, 2 judge, 3 game, 4 jewel, 5 jovial	( )
78.0	*66*	*152*	*20*	*51*	83. mudo	1 silent, 2 mud, 3 much, 4 agitated, 5 angry	( )
86.3	*51*	*107*	*17*	*80*	84. lana	1 side, 2 lake, 3 wool, 4 moon, 5 brace	( )
59.8	*90*	*212*	*31*	*6*	85. resto	1 repose, 2 remainder, 3 peace, 4 stagnation, 5 cure	( )
75.8	*70*	*160*	*15*	*114*	86. sudoeste	1 Southern, 2 Southwest, 3 Southeast, 4 sudorific, 5 Westward	( )
66.3	*83*	*191*	*23*	*27*	87. ejecutar	1 exercise, 2 eject, 3 ejaculate, 4 execute, 5 educate	( )
92.8	*25*	*61*	*10*	*159*	88. caminar	1 camera, 2 camp, 3 picture, 4 whisper, 5 walk	( )
73.5	*73*	*165*	*18*	*69*	89. despacio	1 in front, 2 despatch, 3 straight, 4 slowly, 5 rapidly	( )
73.5	*74*	*166*	*32*	*3*	90. humo	1 man, 2 ground, 3 human, 4 smoke, 5 bone	( )
64.4	*85*	*196*	*17*	*81*	91. carro	1 dear, 2 wagon, 3 affectionate, 4 face, 5 case	( )
62.1	*88*	*204*	*11*	*48*	92. escoba	1 broom, 2 school, 3 beetle, 4 estate, 5 escape	( )
75.8	*71*	*161*	*19*	*56*	93. fondo	1 form, 2 fondness, 3 friendship, 4 experience, 5 background	( )
41.2	*96*	*235*	*46*	*1*	94. propósito	1 probable, 2 intention, 3 proportion, 4 propose, 5 noun	( )
75.5	*72*	*163*	*23*	*26*	95. lograr	1 call, 2 place, 3 seek, 4 succeed in, 5 fill	( )

1	2	3	4	5			
50.3	*92*	*224*	*13*	*130*	96. sugerir	1 to sweeten, 2 believe, 3 endure, 4 suggest, 5 suffer.....	( )
46.1	*94*	*231*	*29*	*10*	97. morder	1 murder, 2 bite, 3 change, 4 hurt, 5 chew.....	( )
21.2	*99*	*261*	*6*	*226*	98. inmediato	1 immaterial, 2 soon, 3 expedient, 4 measured, 5 next.....	( )
24.5	*98*	*257*	*19*	*60*	99. advertir	1 advance, 2 adverse, 3 object, 4 observe, 5 avoid.....	( )
17.3	*100*	*271*	*21*	*41*	100. mudar(se)	1 murmur, 2 to be silent, 3 mock, 4 change, 5 drive.....	( )

PART II — COMPREHENSION

DIRECTIONS: *The following statements are either true or false. Put a plus sign (+) at the right of the statements that are true, and put a zero (0) at the right of the statements that are false. If you do not understand a sentence it is better to omit it rather than guess at its answer. Do all the easy ones first, then return to the harder ones if you have time. Twenty minutes.*

EXAMPLES:

Los guantes se llevan en los pies.....(0)
Las navajas sirven para cortar.....(+)
Nunca hace sol en América.....(0)

Per cent correct answers	*Order of difficulty within parts*	*Order of difficulty whole test*	*Index of goodness*	*Order of goodness whole test*	
1	2	3	4	5	
96.7	*18*	*35*	*5*	*238*	1. Hay más de cuarenta estados en los Estados Unidos.....( )
99.7	*3*	*6*	*4*	*253*	2. Hay nueve días en una semana.....( )
100.0	*1*	*2*	*4*	*254*	3. El hombre tiene dos pies.....( )
98.4	*8*	*22*	*6*	*227*	4. Los hombres son más inteligentes que los animales.....( )
98.4	*9*	*23*	*5*	*245*	5. El español se habla en España.....( )
99.3	*5*	*8*	*5*	*239*	6. Los perros tienen cinco pies.....( )
98.0	*12*	*26*	*6*	*218*	7. La semana tiene diez días.....( )
96.4	*20*	*38*	*4*	*255*	8. Cuba no está en la Argentina.....( )
99.3	*6*	*9*	*4*	*259*	9. La ensalada es una de las estaciones del año.....( )
99.7	*4*	*7*	*5*	*246*	10. El primer mes del año es el lunes.....( )
97.0	*16*	*33*	*6*	*219*	11. La hermana de mi madre es mi tía.....( )
99.0	*7*	*12*	*5*	*247*	12. Madrid es la capital de España.....( )
92.5	*32*	*67*	*8*	*189*	13. Hace más de doscientos años que voy a la escuela.....( )
90.5	*37*	*84*	*7*	*206*	14. El estado de North Dakota está muy cerca del estado de Texas.....( )
98.4	*10*	*24*	*4*	*260*	15. Marzo es el tercer mes del año.....( )

1	2	3	4	5	
98.4	*11*	*25*	*6*	*220*	16. Cuando uno tiene hambre quiere beber agua y no comer.....( )
100.0	*2*	*3*	*4*	*256*	17. Hay muchas cosas bellas en el mundo.....( )
98.0	*13*	*27*	*6*	*228*	18. A algunos muchachos les gusta jugar y correr mucho.....( )
93.4	*29*	*56*	*5*	*240*	19. El gobernador de un estado es siempre la persona más inteligente del estado.( )
96.4	*21*	*39*	*6*	*229*	20. Un perro puede correr cien kilómetros con más velocidad que un tren de ferrocarril.....( )
91.2	*36*	*79*	*8*	*190*	21. El oro vale más que la plata.....( )
95.8	*22*	*41*	*7*	*195*	22. En el verano los días son más calurosos que en el invierno.....( )
94.1	*28*	*53*	*6*	*221*	23. Los mecheros de gas son más cómodos que las luces eléctricas.....( )
92.1	*34*	*70*	*11*	*149*	24. El acento escrito se usa menos en español que en inglés.....( )
98.0	*14*	*38*	*6*	*230*	25. Muchos animales, como el caballo, el perro, la vaca, etc., son muy útiles al hombre.....( )
85.6	*46*	*109*	*7*	*207*	26. Cinco más quince son veinte.....( )
94.8	*24*	*48*	*5*	*241*	27. Algunos árboles pierden las hojas en el otoño.....( )
87.2	*43*	*101*	*7*	*196*	28. El subjuntivo se emplea siempre después del verbo *decir*.....( )
94.8	*25*	*49*	*7*	*197*	29. Hay más preposiciones que verbos en español.....( )
84.0	*51*	*123*	*3*	*263*	30. En las partes centrales de África hace mucho calor en ciertas estaciones....( )
97.7	*15*	*29*	*5*	*242*	31. Hay unos ríos muy grandes en las Américas y un gran número de ríos pequeños( )
95.4	*23*	*43*	*7*	*208*	32. En los Estados Unidos se publican menos de diez periódicos.....( )
85.6	*47*	*110*	*4*	*172*	33. Cuando uno le da a Vd. las gracias por un favor que Vd. le ha hecho, es correcto responder "No hay de que".....( )

1	2	3	4	5	
86.9	*44*	*104*	*2*	*266*	34. Philadelphia es una de las ciudades más grandes de las Américas.....( )
82.3	*55*	*135*	*11*	*145*	35. La estación del frío es el invierno, cuando el sol sale tarde y se pone temprano resultando por ésto muy largas las noches.....( )
97.0	*17*	*34*	*6*	*222*	36. Cuando una persona envía una carta a un amigo suyo que vive en una ciudad muy grande basta poner en el sobre nada más que su nombre.....( )
61.8	*69*	*205*	*19*	*61*	37. Cuando llueve se moja la tierra.....( )
94.4	*27*	*52*	*8*	*183*	38. Hoy los grandes vapores transatlánticos son muy cómodos y seguros.....( )
88.6	*40*	*97*	*11*	*150*	39. Las bebidas que se toman corrientemente con las comidas en los Estados Unidos son el café y el té.....( )
93.0	*31*	*60*	*8*	*178*	40. El caballero bien vestido lleva la corbata y los guantes en el pie izquierdo..( )
88.6	*41*	*98*	*18*	*76*	41. Una madre buena cuida con cariño a su niñito.....( )
85.0	*48*	*113*	*9*	*167*	42. Inglaterra se halla en dos islas que están en el océano pacífico.....( )
96.7	*19*	*36*	*7*	*209*	43. Todos los seres humanos tienen cuatro brazos.....( )
94.8	*26*	*50*	*8*	*184*	44. Cuando uno tiene frío es bueno ponerse más ropa.....( )
92.2	*35*	*71*	*7*	*210*	45. En España hace más calor que en el norte del Canadá.....( )
88.2	*42*	*99*	*7*	*211*	46. El color de una naranja es rojo.....( )
74.2	*62*	*164*	*9*	*173*	47. España es una república.....( )
79.8	*60*	*143*	*15*	*108*	48. Para leer bien hay que aprender de memoria algunas poesías.....( )
93.4	*30*	*57*	*7*	*198*	49. No hay casa en los países hispano-americanos que tenga patio.....( )
90.2	*38*	*87*	*7*	*212*	50. Muchos jóvenes han estudiado el español sin haberlo aprendido bastante bien para charlar efectivamente con sus amigos íntimos.....( )

1	2	3	4	5	
88.9	*39*	*94*	*8*	*179*	51. El verbo *estar* se emplea para expresar las calidades permanentes y esenciales. . ( )
83.6	*52*	*127*	*9*	*168*	52. Si uno se pone con la cara hacia el norte, tiene a la mano izquierda el este y a la derecha el sur. . . . . ( )
73.2	*63*	*167*	*16*	*94*	53. Algunas gallinas ponen huevos. . . . . ( )
79.4	*61*	*149*	*9*	*169*	54. Cuando llegan amigos a nuestra casa para hacernos una visita solemos apresurarnos a darles la bienvenida. . . . . ( )
52.0	*73*	*220*	*12*	*135*	55. La Casa Blanca de Wáshington, D. C., es de más de quinientos metros de altura. . . . . ( )
84.3	*49*	*118*	*16*	*95*	56. Cuando el reloj indica el mediodía o la medianoche, las manecillas están juntas. . . . . ( )
68.0	*66*	*186*	*12*	*136*	57. Cervantes murió el siglo XIX. . . . . ( )
71.6	*65*	*173*	*14*	*121*	58. El artículo indefinido masculino no se usa jamás con una palabra que termina en la letra *a*. . . . . ( )
84.3	*50*	*119*	*16*	*96*	59. Los hermanos no deben jamás ayudarse los unos a los otros. . . . . ( )
92.5	*33*	*68*	*9*	*174*	60. El azúcar es más dulce que la pimienta. . . . . ( )
82.6	*55*	*132*	*17*	*87*	61. Todo hombre que vive tiene cráneo pero algunas personas viven a pesar de haber perdido los ojos. . . . . ( )
85.9	*45*	*108*	*9*	*175*	62. Una muchacha hermosa no es siempre aplicada. . . . . ( )
83.7	*53*	*128*	*3*	*264*	63. La Alhambra es un célebre palacio en Granada. . . . . ( )
80.7	*59*	*141*	*17*	*88*	64. El azúcar es muy amargo. . . . . ( )
47.1	*74*	*229*	*10*	*160*	65. El primer nombre de un mejicano es su apellido. . . . . ( )
72.2	*64*	*172*	*17*	*89*	66. *Hablo* pierde la última vocal cuando precede al sustantivo. . . . . ( )
67.4	*67*	*188*	*4*	*261*	67. Si Vd. mira hacia el oeste, la espalda está al este y el brazo derecho al norte. ( )

1	2	3	4	5	
82.6	*56*	*133*	*15*	*109*	68. El jabón se usa en muchos países cuando uno se baña.....( )
82.0	*58*	*137*	*7*	*199*	69. El hombre que a penas puede correr tres mil metros en treinta minutos debe poder correr seis kilómetros en diez minutos.....( )
67.4	*68*	*189*	*21*	*42*	70. El tocador y la butaca son muebles de la casa.....( )
54.6	*72*	*217*	*19*	*62*	72. El desgraciado nunca echa de menos a sus amigos.....( )
83.0	*54*	*131*	*16*	*97*	72. Las ovejas viven en alta mar.....( )
57.8	*71*	*214*	*17*	*90*	73. Los ladrillos son más duros que el pan.....( )
42.5	*75*	*234*	*15*	*110*	74. Se ha probado que Cristóbal Colón fué el primer europeo que puso pie a tierra en el continente norteamericano.....( )
61.4	*70*	*207*	*28*	*12*	75. "Adelantar" es uno de los sinónimos de "avanzar".....( )

PART III — GRAMMAR

DIRECTIONS: *Each of the English sentences below is followed by a translation which is correct as far as it goes, but which is incomplete. Each translation can be completed correctly by inserting one, two or three words in each blank (——). When you have decided what is necessary to complete the translation, write it on the line at the right of the translation. WRITE LEGIBLY. Do not spend too much time on any one sentence.*

Write carefully and clearly, putting your answers on the lines at the right, as in the examples. Forty-five minutes.

EXAMPLES:

1. What is your name? (——) se llama Vd? ..... ¿Cómo
2. These books are mine. Estos libros (——) ..... son míos
3. Which pen have you? (——) pluma tiene Vd? ..... ¿Cuál

Per cent correct answers	*Order of difficulty within parts*	*Order of difficulty whole test*	*Index of goodness*	*Order of goodness whole test*	
1	2	3	4	5	
84.6	*14*	*115*	*6*	*231*	1. What do you smoke? ¿(——) fuma Vd? .....
93.4	*5*	*58*	*9*	*170*	2. Has John studied his lesson? ¿Ha (——) Juan su lección?
95.1	*1*	*46*	*7*	*200*	3. There are five fingers on each hand. (——) cinco dedos en cada mano .....
95.1	*2*	*47*	*8*	*185*	4. He studies a great deal. Estudia (——) .....
94.8	*3*	*51*	*8*	*191*	5. It is warm, isn't it? (——) calor, ¿No es verdad? .....
92.8	*6*	*62*	*1*	*270*	6. I see the book. Veo (——) .....
92.8	*7*	*63*	*7*	*213*	7. I am here. Estoy (——) .....
91.2	*9*	*80*	*6*	*232*	8. I have not said anything to you. No (——) dicho nada a Vds .....

1	2	3	4	5		
83.3	*17*	*129*	*6*	*223*	9. He doesn't give it to him. No (——) da.....	.....
93.8	*4*	*54*	*6*	*233*	10. Has John any brothers? (——) hermanos Juan?.....	.....
84.3	*15*	*120*	*11*	*151*	11. He had no money. No (——) dinero.....	.....
68.6	*38*	*183*	*13*	*131*	12. How old are you? ¿(——) tiene Vd?.....	.....
89.6	*12*	*91*	*8*	*186*	13. I have seen all. He (——) todo.....	.....
39.5	*66*	*237*	*12*	*143*	14. Hundreds of enemies are in the city. Centenares de enemigos (——) ciudad.....	.....
85.0	*13*	*114*	*16*	*106*	15. I have a good horse. Tengo (——) caballo.....	.....
81.7	*18*	*138*	*16*	*98*	16. The roses are for you. Las rosas son (——).....	.....
68.6	*39*	*184*	*3*	*265*	17. Which book is John's? This one. ¿Cuál libro es de Juan? (——).....	.....
78.4	*25*	*151*	*9*	*176*	18. This is my good friend. Este es (——) amigo.....	.....
92.8	*8*	*64*	*10*	*161*	19. I don't like this town. (——) gusta este pueblo.....	.....
51.6	*58*	*221*	*17*	*82*	20. I have not seen him yet. Ne le he visto (——).....	.....
81.4	*19*	*64*	*17*	*83*	21. Every day I study the lesson. Todos (——) estudio la lección.....	.....

1	2	3	4	5		
62.7	*48*	*202*	*22*	*36*	22. I prefer to study Spanish. (——) el español.....	.....
90.5	*11*	*85*	*11*	*152*	23. We study in order to learn. Estudiamos (——) aprender	.....
70.9	*34*	*177*	*13*	*132*	24. How many girls are there in this class? (——) muchachas hay en esta clase?.....	.....
66.3	*41*	*192*	*22*	*37*	25. This is a wooden table. Esta es una mesa (——).....	.....
63.7	*45*	*199*	*11*	*146*	26. John is in Guatemala. Juan (——) Guatemala.....	.....
84.3	*16*	*121*	*15*	*111*	27. The man and his wife are Italians. El hombre y su esposa (——).....	.....
65.0	*43*	*195*	*8*	*180*	28. If he has time he comes to see me every Sunday. Si (——) el tiempo, viene a visitarme todos los domingos..	.....
78.1	*26*	*153*	*13*	*133*	29. The program is here. El programa (——).....	.....
71.2	*33*	*176*	*17*	*84*	30. Saturday is the first day of the month. El sábado es el (——) del mes.....	.....
79.7	*20*	*144*	*2*	*267*	31. Two days ago I went to the house. (——) dos días fuí a la casa.....	.....
28.7	*81*	*252*	*22*	*33*	32. His head aches. (——) la cabeza.....	.....
69.3	*36*	*181*	*13*	*129*	33. This is the prettiest. Este es (——) hermoso.....	.....
76.4	*29*	*158*	*18*	*70*	34. I studied the art of singing. Estudié el arte (——)....	.....

1	2	3	4	5		
76.8	*28*	*156*	*19*	*63*	35. I don't understand you. No (——) comprendo a Vd..	
24.2	*85*	*258*	*1*	*273*	36. I do not speak Spanish. No (——) castellano.....	
77.1	*27*	*155*	*42*	*2*	37. Spanish is spoken here. Aquí (——) español.....	
57.5	*54*	*215*	*18*	*77*	38. When the youth had finished his courses he left for Spain. Cuando el joven (——) sus cursos partió para España.....	
71.6	*32*	*174*	*18*	*71*	39. I came the first days of the year. Vine (——) días del año.....	
60.8	*51*	*210*	*21*	*43*	40. The boy goes to school singing. El muchacho (——) a la escuela.....	
78.8	*24*	*150*	*15*	*112*	41. I have little money. Tengo (——) diero.....	
75.8	*30*	*162*	*18*	*72*	42. Neither do I want to study. (——) quiero estudiar...	
61.1	*50*	*209*	*14*	*122*	43. I shall not go out lest I catch a cold. (——) por miedo de que coja un resfriado.....	
79.7	*21*	*145*	*18*	*73*	44. I never smoke. (——) fumo.....	
51.6	*59*	*222*	*14*	*123*	45. I sold my black horse yesterday. Ayer (——) caballo negro.....	
69.3	*37*	*182*	*16*	*101*	46. I have not worked for three months. (——) meses que no trabajo.....	

1	2	3	4	5	
20.9	*88*	*262*	*30*	*9*	47. She does not know how much he loves her. Ella no sabe (——) la quiere .....
36.9	*68*	*239*	*18*	*51*	48. Which one of the books do you want? (——) los libros quiere Vd? .....
24.8	*84*	*256*	*22*	*38*	49. Is the lesson difficult? Yes, it is. ¿Es difícil la lección? Sí, (——) es .....
31.4	*75*	*246*	*20*	*52*	50. The more one studies, the more one learns. Cuanto más uno estudia (——) aprende .....
63.0	*47*	*201*	*21*	*44*	51. I found her playing the piano. La hallé (——) el piano.
41.2	*65*	*236*	*19*	*57*	52. Fanny is the younger of the two. Paquita es (——) de las dos .....
61.4	*49*	*208*	*14*	*124*	53. This one is as good as that one. Este es (——) como aquél .....
90.9	*10*	*83*	*2*	*269*	54. Horses and cows are useful animals. Los caballos y las vacas (——) útiles .....
56.8	*55*	*216*	*8*	*181*	55. He wishes to give them to her. Desea (——) a ella....
60.1	*52*	*211*	*19*	*58*	56. We study in order to be able to speak. Estudiamos para (——) .....
46.7	*63*	*230*	*14*	*116*	57. They arrived on Sunday. Llegaron (——) .....
70.0	*35*	*156*	*20*	*53*	58. I read the drama of Shakespeare. Leo (——) drama de Shakespeare .....
64.0	*44*	*197*	*21*	*45*	59. It began to rain early. Empezó (——) temprano .....

1	2	3	4	5		
9.5	*99*	*274*	*10*	*156*	60. For that (purpose) it is necessary to work a great deal. (——) es preciso trabajar mucho	
59.1	*53*	*213*	*15*	*113*	61. He died from fear. Murió (——)	
53.0	*57*	*219*	*23*	*28*	62. My mother wanted me to go to the university this year. Mi madre quería que yo (——) a la universidad este año	
79.7	*22*	*146*	*23*	*29*	63. Hardly had he left his home when it began to rain. Apenas hubo salido de su casa, cuando (——) llover	
65.3	*42*	*194*	*21*	*47*	64. No one has seen me. (——) ha visto	
72.9	*31*	*170*	*25*	*20*	65. I have already written the letter. (——) escrito la carta	
63.7	*46*	*200*	*22*	*39*	66. Before translating let us read the Spanish. Antes (——) leamos el español	
49.0	*60*	*225*	*32*	*4*	67. The book was John's. El libro (——) Juan	
29.7	*79*	*250*	*25*	*21*	68. I should like to see her once more. (——) otra vez	
30.7	*76*	*247*	*18*	*74*	69. From the mountain we saw Quito. Desde la montaña (——) Quito	
35.6	*71*	*242*	*10*	*157*	70. I shall do it for my mother's sake. Lo haré (——) madre	
79.8	*23*	*147*	*22*	*34*	71. I think so. Creo (——) sí	
29.4	*80*	*251*	*21*	*46*	72. It may be possible that he has not received our letter. Puede que no (——) nuestra carta	

1	2	3	4	5		
68.3	*40*	*185*	*24*	*22*	73. I forbid your doing it. Prohibo que Vd. (——).....	.....
35.6	*72*	*243*	*32*	*5*	74. You cannot learn a foreign language unless you study hard. No puede Vd. aprender una lengua extranjera a menos (——) con ahinco.....	.....
43.8	*64*	*233*	*23*	*30*	75. Whatever you do, sir, do it well. Cualquiera cosa que haga, señor, (——) bien.....	.....
27.8	*82*	*253*	*14*	*125*	76. Which horse has John? Ours. ¿Cuál caballo tiene Juan? (——).....	.....
36.3	*70*	*241*	*16*	*107*	77. This room is fifteen meters wide. Este cuarto (——) metros de ancho.....	.....
20.9	*89*	*263*	*19*	*59*	78. It is impossible for me to do it. (——) imposible hacerlo	.....
20.9	*90*	*264*	*19*	*64*	79. Do you know anybody here? ¿Conoce Vd. (——) aquí?	.....
48.0	*62*	*228*	*16*	*102*	80. For the next time study the next lesson. (——) próxima vez estudie la lección siguiente.....	.....
20.6	*92*	*266*	*16*	*103*	81. He would not do it. No (——) hacerlo.....	.....
30.4	*77*	*248*	*31*	*7*	82. Are you an American? I am. ¿Es Vd. americano? (——) soy.....	.....
30.4	*78*	*249*	*23*	*31*	83. Introduce me to her. (——) Vd. a ella.....	.....
27.1	*83*	*254*	*23*	*32*	84. He dressed in black. Se (——) negro.....	.....

1	2	3	4	5		
49.0	*61*	*226*	*17*	*63*	85. I asked the teacher himself. Le pregunté al maestro (——).....	.....
36.6	*69*	*240*	*28*	*13*	86. It is to be hoped. Es (——) esperar.....	.....
39.5	*67*	*238*	*24*	*24*	87. May the Lord help me! ¡Que Dios (——)!.....	.....
12.7	*98*	*273*	*10*	*162*	88. He was here a whole day. Estuvo aquí (——) día....	.....
19.9	*94*	*268*	*7*	*201*	89. Yesterday I told you to know your lesson every day. Ayer les he dicho a Vds. que (——) su lección cada día	.....
15.4	*97*	*272*	*12*	*137*	90. I am in style. Estoy (——).....	.....
7.2	*100*	*275*	*12*	*138*	91. How many friends! ¡Qué (——) amigos!.....	.....
20.9	*91*	*265*	*13*	*134*	92. What is the price? ¿(——) se vende?.....	.....
22.2	*87*	*260*	*10*	*163*	93. These sell at one peseta and a half a pound. Estos se venden a peseta y media (——).....	.....
22.9	*86*	*259*	*17*	*92*	94. You don't know how pretty Carmen is. Vd. no sabe (——) que es Carmen.....	.....
53.2	*56*	*218*	*10*	*164*	95. He had just arrived when I saw him. (——) llegar cuando le ví.....	.....
18.0	*96*	*270*	*1*	*271*	96. Come here. Venga Vd. (——).....	.....
20.3	*93*	*267*	*18*	*75*	97. Any one may do that. (——) puede hacer eso.....	.....

1	2	3	4	5	
19.0	*95*	*269*	*14*	*117*	98. He cut his finger. El (——) el dedo..... ----------
33.7	*73*	*244*	*31*	*8*	99. This should not disturb us. Esto no debe (——)..... ----------
31.7	*74*	*245*	*19*	*65*	100. If I were to pass my examinations my parents would be glad. Si yo (——) aprobado en los exámenes, mis padres estarían contentos..... ----------

REGENTS EXAMINATION IN GERMAN FOR JUNE, 1925

SECOND HALF

2nd, 3rd and 4th Year Classes

PART I — VOCABULARY

DIRECTIONS: *In each line select the English word or phrase corresponding to the German word at the left, and put its number on the line at the right, as in the examples. Do all the easy ones first. Twenty-five minutes.*

EXAMPLES:

1. Kraft 1 craft, 2 crossing, 3 strap, 4 stream, 5 strength.....(5)
2. laufen 1 laugh, 2 run, 3 let, 4 rest, 5 love.....(2)
3. Schlacht 1 shuttle, 2 balance, 3 settler, 4 battle, 5 search.....(4)

Per cent correct answers	*Order of difficulty within parts*	*Order of difficulty whole test*	*Index of goodness*	*Order of goodness whole test*			
1	2	3	4	5			
98.7	*11*	*18*	*5*	*220*	1. neben	1 beside, 2 bother, 3 nibble, 4 upside down, 5 never.....( )
96.8	*24*	*38*	*−1*	*269*	2. abends	1 in the evening, 2 however, 3 finally, 4 breathing, 5 endless..( )
97.7	*20*	*29*	*2*	*248*	3. zweimal	1 twice, 2 doubt, 3 silent, 4 even, 5 divisible.....( )
98.4	*14*	*22*	*2*	*249*	4. Luft	1 lift, 2 look, 3 laugh, 4 art, 5 air.....( )
99.0	*5*	*12*	*4*	*231*	5. klar	1 colored, 2 clear, 3 rare, 4 classic, 5 cloudy.....( )
98.1	*18*	*27*	*4*	*232*	6. pfeifen	1 whistle, 2 season, 3 jump, 4 ring, 5 reduce.....( )
87.0	*57*	*102*	*9*	*190*	7. fliehen	1 flow, 2 follow, 3 fall, 4 fill, 5 flee.....( )
97.7	*21*	*30*	*7*	*207*	8. richtig	1 bold, 2 wrong, 3 doubtful, 4 right, 5 rash.....( )
96.1	*28*	*47*	*2*	*250*	9. schreien	1 cry, 2 shrive, 3 write, 4 whisper, 5 divide.....( )
98.7	*12*	*19*	*2*	*251*	10. ruhig	1 quiet, 2 queer, 3 ruddy, 4 rueful, 5 rowdy.....( )

1	2	3	4	5		
91.9	*39*	*72*	*13*	*162*	11. verbrennen	1 burn, 2 glimmer, 3 race, 4 forbear, 5 forbid.....( )
99.7	*1*	*4*	*3*	*239*	12. Liebe	1 love, 2 inference, 3 life, 4 remainder, 5 lapse.....( )
98.4	*15*	*23*	*5*	*221*	13. Jugend	1 youth, 2 yoke, 3 mood, 4 virtue, 5 smiling.....( )
99.0	*6*	*13*	*−1*	*270*	14. anstatt	1 instead of, 2 municipal, 3 stated, 4 besides, 5 stately.....( )
99.0	*7*	*14*	*4*	*233*	15. Freundschaft	1 friendship, 2 desertion, 3 shaft, 4 animosity, 5 loyalty.....( )
93.2	*37*	*65*	*2*	*252*	16. nass	1 curious, 2 dry, 3 rash, 4 wasteful, 5 wet.....( )
99.0	*8*	*15*	*3*	*240*	17. Fehler	1 relative, 2 feeler, 3 fellow, 4 cousin, 5 error.....( )
99.4	*3*	*9*	*4*	*234*	18. Bleistift	1 blindspot, 2 comb, 3 blister, 4 fountain pen, 5 pencil.....( )
96.8	*25*	*39*	*10*	*184*	19. Freiheit	1 fright, 2 freshness, 3 fray, 4 freedom, 5 freight.....( )
95.4	*32*	*52*	*9*	*192*	20. gesund	1 harmonious, 2 sinful, 3 gesticulated, 4 unsound, 5 healthy...( )
99.4	*4*	*10*	*4*	*235*	21. weiter	1 further, 2 open, 3 served, 4 wicked, 5 moist.....( )
87.7	*55*	*99*	*22*	*107*	22. abwischen	1 wish, 2 waste, 3 wipe off, 4 send away, 5 separate.....( )
97.7	*22*	*31*	*2*	*253*	23. Wurst	1 worth, 2 sausage, 3 wolf, 4 thirst, 5 burst.....( )
96.4	*26*	*42*	*9*	*192*	24. erwarten	1 accompany, 2 await, 3 take care of, 4 heat, 5 cancel.....( )
99.7	*2*	*5*	*3*	*241*	25. plötzlich	1 in place of, 2 suddenly, 3 fully, 4 obviously, 5 shy.....( )
89.6	*50*	*90*	*17*	*136*	26. reissen	1 sweep, 2 bite, 3 buy, 4 tear, 5 praise.....( )
98.4	*16*	*24*	*4*	*236*	27. wiedersehen	1 overlook, 2 concentrate, 3 squint, 4 see again, 5 be shortsighted( )
90.0	*48*	*87*	*25*	*90*	28. schliessen	1 close, 2 cultivate, 3 open, 4 show, 5 slice.....( )
91.9	*40*	*73*	*19*	*127*	29. heilig	1 hardy, 2 healed, 3 blasphemous, 4 holy, 5 whole.....( )
99.0	*9*	*16*	*4*	*237*	30. Dummheit	1 dimness, 2 imprecation, 3 stupidity, 4 donation, 5 duplication ( )

1	2	3	4	5			
93.5	*35*	*63*	*3*	*242*	31. müde	1 nude, 2 moody, 3 tough, 4 tired, 5 tried	( )
99.0	*10*	*17*	*3*	*243*	32. Gebirge	1 mountain range, 2 pear, 3 castle, 4 gift, 5 bear-den	( )
93.5	*36*	*64*	*12*	*168*	33. heutig	1 to-day's, 2 skinny, 3 leather, 4 joyful, 5 heady	( )
91.0	*45*	*82*	*16*	*142*	34. Geschenk	1 sham, 2 present, 3 permission, 4 shadow, 5 powder	( )
91.3	*42*	*78*	*21*	*114*	35. kostbar	1 precious, 2 counterfeit, 3 coastal, 4 bareheaded, 5 cheap	( )
94.8	*34*	*59*	*10*	*185*	36. beissen	1 bind, 2 bisect, 3 bite, 4 besides, 5 besiege	( )
80.6	*67*	*125*	*45*	*18*	37. bequem	1 beckoning, 2 comfortable, 3 cautious, 4 calm, 5 querulous	( )
97.4	*23*	*33*	*7*	*208*	38. Besuch	1 visit, 2 beseeching, 3 search, 4 trial, 5 sigh	( )
98.1	*19*	*28*	*5*	*222*	39. Tannenbaum	1 coat of tan, 2 fir tree, 3 oak, 4 shrub, 5 trunk	( )
96.1	*29*	*48*	*7*	*209*	40. klug	1 cloudy, 2 clogged, 3 close, 4 clever, 5 clear	( )
90.3	*46*	*83*	*20*	*120*	41. erschrecken	1 frighten, 2 shrug, 3 mistake, 4 escape, 5 shriek	( )
98.7	*13*	*20*	*4*	*238*	42. fortgehen	1 leave, 2 climb, 3 fasten, 4 fortify, 5 increase	( )
74.8	*77*	*155*	*54*	*6*	43. Gefahr	1 draught, 2 drainage, 3 danger, 4 daring, 4 donation	( )
95.4	*33*	*53*	*8*	*203*	44. Neffe	1 uncle, 2 negative, 3 net, 4 nephew, 5 neighbor	( )
96.1	*30*	*49*	*2*	*254*	45. backen	1 back, 2 bicker, 3 beg, 4 bake, 5 cut	( )
85.8	*60*	*108*	*24*	*94*	46. angenehm	1 eager, 2 agreeable, 3 robust, 4 anonymous, 5 rosy	( )
83.2	*61*	*117*	*21*	*115*	47. beschreiben	1 bewail, 2 bewitch, 3 bestow, 4 describe, 5 prescribe	( )
96.4	*27*	*43*	*9*	*193*	48. vorbeifahren	1 linger, 2 ride past, 3 forget, 4 reappear, 5 clear	( )
79.7	*71*	*135*	*32*	*55*	49. leiden	1 loath, 2 lead, 3 tremble, 4 suffer, 5 punish	( )
91.9	*41*	*74*	*16*	*143*	50. draussen	1 out-of-doors, 2 draft, 3 destroy, 4 inside, 5 aroused	( )
90.0	*49*	*88*	*12*	*169*	51. Schatten	1 shadow, 2 shad, 3 shabbiness, 4 shallowness, 5 ferns	( )
82.2	*63*	*120*	*31*	*62*	52. schütteln	1 shuttle, 2 shake, 3 shower, 4 break, 5 burden	( )

1	2	3	4	5			
78.7	*75*	*141*	*33*	*47*	53. aufmerksam	1 attentive, 2 upright, 3 uproarious, 4 remarkable, 5 peculiar.	( )
90.3	*47*	*84*	*16*	*144*	54. Hütte	1 hush, 2 hunt, 3 hut, 4 skin, 5 hospital.	( )
96.1	*31*	*50*	*1*	*261*	55. Arzt	1 art, 2 answer, 3 doorstep, 4 physician, 5 abbot	( )
88.0	*52*	*95*	*27*	*78*	56. Schuld	1 guilt, 2 school, 3 shuttle, 4 shield, 5 shelter	( )
98.4	*17*	*25*	*6*	*214*	57. Gebäude	1 custom, 2 giver, 3 alms, 4 building, 5 mason	( )
67.7	*87*	*182*	*54*	*7*	58. retten	1 soften, 2 ride, 3 save, 4 shimmer, 5 thrive	( )
91.3	*43*	*79*	*22*	*108*	59. Beispiel	1 toy, 2 speech, 3 example, 4 playground, 5 playwright	( )
82.6	*62*	*119*	*39*	*27*	60. nötig	1 necessary, 2 knotty, 3 notably, 4 noticeable, 5 notorious	( )
79.3	*73*	*137*	*31*	*63*	61. kämpfen	1 camp, 2 fight, 3 favor, 4 comb, 5 cluster	( )
81.3	*66*	*124*	*22*	*109*	62. stumm	1 mute, 2 stormy, 3 voiced, 4 stumbling, 5 stupid	( )
79.7	*72*	*136*	*29*	*74*	63. Zweig	1 silence, 2 twig, 3 indication, 4 wisdom, 5 pair	( )
60.3	*93*	*202*	*29*	*75*	64. beklagen	1 claim, 2 call, 3 laud, 4 lament, 5 clatter	( )
82.2	*64*	*121*	*45*	*19*	65. versammeln	1 versify, 2 reverse, 3 forsake, 4 assemble, 5 trouble	( )
66.4	*89*	*189*	*33*	*48*	66. Brille	1 glasses, 2 cricket, 3 edge, 4 scissors, 5 needle.	( )
73.2	*80*	*165*	*38*	*29*	67. drinnen	1 drill, 2 drain, 3 inside, 4 enter, 5 forward.	( )
71.6	*81*	*170*	*32*	*56*	68. jedoch	1 jealous, 2 each other, 3 because, 4 however, 5 just	( )
86.7	*59*	*104*	*1*	*262*	69. Tasse	1 tassel, 2 tissue, 3 cup, 4 saucer, 5 tap	( )
93.2	*38*	*66*	*5*	*223*	70. Maler	1 miller, 2 painter, 3 druggist, 4 grocer, 5 manager	( )
88.0	*53*	*96*	*23*	*101*	71. gebrauchen	1 bring, 2 marry, 3 use, 4 smoke, 5 storm	( )
91.3	*44*	*80*	*12*	*170*	72. Ofen	1 open spaces, 2 oven, 3 heat, 4 east, 5 door.	( )
74.8	*78*	*156*	*36*	*36*	73. höflich	1 hopeful, 2 polite, 3 homelike, 4 positive, 5 pertinent	( )

1	2	3	4	5			
76.1	*76*	*150*	*16*	*145*	74.	besehen	1 imagine, 2 beseech, 3 to look at, 4 visit, 5 mistake.....( )
80.6	*68*	*126*	*18*	*131*	75.	Weile	1 reason, 2 while, 3 midst, 4 wave, 5 oak.....( )
88.0	*54*	*97*	*23*	*102*	76.	Schlittschuh	1 shoe-black, 2 boot, 3 glove, 4 sole, 5 skate.....( )
79.0	*74*	*139*	*31*	*64*	77.	bestrafen	1 stray, 2 betray, 3 punish, 4 command, 5 counsel.....( )
87.7	*56*	*100*	*9*	*194*	78.	obwohl	1 obviously, 2 healthy, 3 frequently, 4 obligatory, 5 although..( )
74.8	*79*	*157*	*58*	*4*	79.	Mahlzeit	1 mill, 2 meal, 3 multiplication, 4 generation, 5 recipe.....( )
80.3	*70*	*130*	*38*	*30*	80.	beeilen	1 bewail, 2 bail, 3 retard, 4 boil, 5 hurry.....( )
89.0	*51*	*93*	*32*	*57*	81.	sonderbar	1 without funds, 2 peculiar, 3 beardless, 4 sunny, 5 cloudy....( )
71.3	*82*	*172*	*50*	*11*	82.	sogar	1 earnest, 2 so-called, 3 sugary, 4 savory, 5 even.....( )
58.7	*95*	*208*	*47*	*14*	83.	rauschen	1 rustle, 2 smoke, 3 outside, 4 burn, 5 dream.....( )
65.4	*92*	*194*	*21*	*116*	84.	erhalten	1 contain, 2 receive, 3 overhaul, 4 grow cold, 5 behold.....( )
68.7	*86*	*180*	*24*	*95*	85.	buchstabieren	1 classify, 2 spell, 3 break, 4 stabilize, 5 read proof.....( )
82.2	*65*	*122*	*17*	*137*	86.	schmecken	1 smile, 2 taste, 3 smell, 4 smoulder, 5 smelt.....( )
69.7	*85*	*178*	*36*	*37*	87.	bewegen	1 move, 2 because, 3 weigh, 4 interrupt, 5 settle.....( )
70.0	*84*	*176*	*37*	*33*	88.	dennoch	1 nightly, 2 also, 3 nevertheless, 4 afterwards, 5 denoting.....( )
66.4	*90*	*189*	*40*	*24*	90.	Sieg	1 sickness, 2 failure, 3 victory, 4 identify, 5 scythe.....( )
70.6	*83*	*173*	*32*	*58*	90.	hässlich	1 hasty, 2 ugly, 3 haughty, 4 domestic, 5 healthy.....( )
66.4	*91*	*190*	*22*	*110*	91.	Kranz	1 wreath, 2 crown, 3 craving, 4 whim, 5 cruelty.....( )
51.9	*99*	*221*	*47*	*15*	92.	dicht	1 wide, 2 composed, 3 tough, 4 dexterous, 5 close.....( )
57.1	*97*	*212*	*59*	*3*	93.	Mitglied	1 member, 2 withdrawal, 3 reduction, 4 sympathy, 5 addition.( )
52.9	*98*	*219*	*48*	*13*	94.	begreifen	1 grieve, 2 grope, 3 beg, 4 pray, 5 understand.....( )
22.3	*100*	*265*	*1*	*263*	95.	einfallen	1 invade, 2 infer, 3 tumble, 4 excel, 5 take in.....( )

1	2	3	4	5		
80.6	*69*	*127*	*35*	*38*	96. ungefähr	1 unfavorable, 2 travelled, 3 about, 4 ungainly, 5 safe ..... ( )
60.3	*94*	*203*	*51*	*9*	97. Speise	1 glasses, 2 sparrow, 3 food, 4 pride, 5 spice ..... ( )
67.7	*88*	*183*	*33*	*49*	98. vorüber	1 above, 2 past, 3 forth, 4 further, 5 forsaken..... ( )
87.0	*58*	*103*	*30*	*68*	99. dünn	1 poetic, 2 close, 3 thin, 4 thoughtful, 5 fat..... ( )
57.4	*96*	*211*	*78*	*1*	100. Kahn	1 corn, 2 calmness, 3 boat, 4 classic, 5 cloudy..... ( )

Per cent correct answers	*Order of difficulty within parts*	*Order of difficulty whole test*	*Index of goodness*	*Order of goodness whole test*
1	2	3	4	5
99.7	*4*	*6*	*2*	*255*
99.7	*5*	*7*	*3*	*244*
97.4	*11*	*34*	*6*	*215*
99.7	*6*	*8*	*3*	*245*
96.5	*15*	*44*	*7*	*210*
98.4	*9*	*26*	*5*	*224*
98.7	*8*	*21*	*3*	*246*
97.1	*12*	*36*	*−2*	*272*
99.4	*7*	*11*	*2*	*256*
95.5	*19*	*54*	*5*	*225*
96.5	*16*	*45*	*10*	*186*
96.8	*14*	*40*	*6*	*216*
95.2	*20*	*56*	*7*	*211*
95.2	*21*	*57*	*13*	*163*
96.5	*17*	*46*	*11*	*175*

PART II — COMPREHENSION

DIRECTIONS: *The following statements are either true or false. Put a plus sign (+) at the right of the statements that are true, and put a zero (0) at the right of the statements that are false. If you do not understand a sentence, it is better to omit it than to guess at its answer. Do all the easy ones first, then return to the harder ones if you have time. Twenty minutes.*

EXAMPLES:

1. In Afrika gibt es noch wilde Tiere... .....(+)
2. Nach dem Nordpol fahrt man mit der Eisenbahn.....(0)
3. In den Volkschulen werden viele Kinder unterrichtet....(+)

1. Schwarz ist weiss.....( )
2. Kalt ist heiss.....( )
3. Jede Familie besteht aus sechs Personen.....( )
4. Drei mal drei ist zehn.....( )
5. Es gibt keine Menschen mit blonden Haaren.....( )
6. Schuhe trägt man an den Händen, Handschuhe an den Füssen.....( )
7. In Deutschland spricht man Deutsch.....( )
8. Im Sommer haben die Tage sechs und zwanzig Stunden.....( )
9. Jedes Jahr gehen viele Kinder zur Schule.....( )
10. Wir haben zehn Finger an jeder Hand.....( )
11. Die meisten Menschen werden über hundert Jahre alt.....( )
12. Alle Schüler sind immer fleissig.....( )
13. Oben auf dem hohen Berge ist es immer wärmer als unten im Tal.....( )
14. Kurz bedeutet lang.....( )
15. Ein Viertel bedeutet mehr als die Hälfte.....( )

1	2	3	4	5	
97.1	*13*	*37*	*6*	*217*	16. Im Winter trägt man gewöhnlich leichtere Kleider als im Sommer.....( )
97.7	*10*	*32*	*5*	*226*	17. Frühstück wird überall um zwölf Uhr nachts gegessen.....( )
100.0	*1*	*1*	*2*	*257*	18. Das ganze englische Alphabet besteht aus zehn Buchstaben.....( )
95.8	*18*	*51*	*6*	*218*	19. Der Plural von *der Mann* heisst *die Männer*.....( )
94.6	*25*	*62*	*−1*	*271*	20. Nache einem heissen Tage hofft man auf eine kühle Nacht.....( )
92.0	*30*	*75*	*12*	*171*	21. Gold ist wertvoller als Silber.....( )
100.0	*2*	*2*	*2*	*258*	22. Wer blind ist, kann nicht sehen.....( )
92.6	*28*	*70*	*10*	*187*	23. In Amerika gibt es viele grosse aber auch viele kleine Flüsse und Seen.....( )
100.0	*3*	*3*	*2*	*259*	24. Von New York nach England fährt man mit der Eisenbahn.....( )
95.2	*22*	*58*	*1*	*264*	25. Wenn etwas zu gross ist, dann ist es nicht gross genug.....( )
83.9	*42*	*115*	*16*	*146*	26. Walfische leben im Ozean.....( )
91.6	*31*	*76*	*8*	*204*	27. Ein Mann, der nicht weit sehen kann, ist kurzsichtig.....( )
90.4	*32*	*85*	*14*	*157*	28. Brot bäckt man aus Mehl.....( )
94.8	*23*	*60*	*11*	*186*	29. Beim Essen sitzt man gewöhnlich auf dem Fussboden.....( )
89.4	*34*	*92*	*7*	*212*	30. Alle Wolken bringen Regen.....( )
85.8	*39*	*109*	*14*	*158*	31. Was man einmal gelesen hat, vergisst man nie wieder.....( )
92.2	*29*	*71*	*9*	*195*	32. Das Gegenteil von schuldig ist unschuldig.....( )
80.7	*45*	*128*	*27*	*79*	33. Nichts wissen ist schlimm, nichts wissen wollen schlimmer.....( )
94.8	*24*	*61*	*11*	*177*	34. Gebäude haben nie mehr als drei Stockwerke.....( )
88.4	*35*	*94*	*17*	*138*	35. Der Sommer ist eine Jahreszeit.....( )
80.7	*46*	*129*	*30*	*69*	36. Man soll sich nie entschuldigen, wenn man jemandem auf die Füsse tritt....( )
80.0	*49*	*134*	*18*	*132*	37. Es ist leicht, es allen Menschen recht zu machen.....( )

1	2	3	4	5	
86.4	*38*	*107*	*18*	*133*	38. Wenige Menschen wissen, wie viel man wissen muss, um zu wissen, wie wenig man weiss.....( )
88.0	*36*	*98*	*11*	*178*	39. Es ist oft sehr schwer, die Gedanken eines anderen zu erraten.....( )
78.4	*52*	*144*	*20*	*121*	40. Scheint die Sonne noch so schön, einmal muss sie untergehn.....( )
89.7	*33*	*91*	*5*	*227*	41. Morgen ist nicht heute.....( )
78.4	*53*	*145*	*17*	*139*	42. Gute Eltern versuchen für ihre Kinder zu sorgen.....( )
80.3	*47*	*131*	*24*	*96*	43. Versprechen ist leichter als Wort halten.....( )
80.3	*48*	*132*	*31*	*65*	44. Je deutlicher ein Mensch spricht, je schwerer kann man ihn verstehen.....( )
93.3	*26*	*67*	*6*	*219*	45. Der Januar ist der kürzeste Monat im Jahre.....( )
85.4	*40*	*110*	*11*	*179*	46. Im Herbst verlieren manche Bäume ihre Blätter.....( )
83.6	*43*	*116*	*23*	*103*	47. Ein Hügel ist höher als ein Berg.....( )
74.8	*57*	*158*	*31*	*66*	48. Bienen und Ameisen sind fleissige Insekten.....( )
86.8	*37*	*105*	*15*	*153*	49. Gute Maler, Musiker und Architekten sind Künstler.....( )
93.3	*27*	*68*	*16*	*147*	50. Zu Weihnachten werden viele Tannenbäume verkauft.....( )
69.7	*64*	*179*	*7*	*213*	51. Wenn die Sonne scheint, kann man leicht über seinen Schatten springen.....( )
19.3	*75*	*269*	*−8*	*275*	52. Die Sonne bewegt sich um die Erde und die Erde um den Mond.....( )
79.4	*50*	*138*	*−5*	*274*	53. Manche Hennen legen Eier.....( )
75.8	*55*	*152*	*38*	*31*	54. Der Adler ist kleiner als die Taube.....( )
78.7	*51*	*142*	*34*	*44*	55. Der Krieg zerstört, der Frieden baut auf, sagt ein deutsches Sprichwort.....( )
69.9	*60*	*162*	*43*	*21*	56. Das amerikanische Silbergeld ist viereckig.....( )
57.1	*69*	*213*	*16*	*148*	57. Es ist bewiesen, dass der Mensch zufrieden wäre, wenn er all das Geld hätte, das er sich wünschen kann.....( )

1	2	3	4	5	
84.2	*41*	*114*	*20*	*122*	58. Ein höflicher Mann kommt leichter durch die Welt als ein Grobian.....( )
74.2	*59*	*161*	*12*	*172*	59. Wenn der Mensch etwas nicht tun will, sagt er zuweilen: "ich kann nicht," und wenn er etwas nicht kann, sagt er oft: "ich will nicht."..... ( )
69.9	*61*	*163*	*40*	*25*	60. Im Nebel kann man nicht weit sehen..... ( )
75.8	*56*	*153*	*38*	*32*	61. Wer viel gereist ist, hat manches gesehen, ausser wenn er blind ist..... ( )
83.3	*44*	*118*	*21*	*117*	62. Wenn Uhren aufgezogen sind, laufen sie für immer..... ( )
74.5	*58*	*159*	*14*	*159*	63. Jede Dummheit findet einen, der sie macht..... ( )
45.5	*73*	*231*	*2*	*260*	64. Wer in die strahlende Sonne schaut, wird blind..... ( )
67.8	*65*	*184*	*39*	*28*	65. Viele Menschen leiden an Geldmangel..... ( )
72.6	*62*	*166*	*26*	*84*	66. Alle Kinder lernen in kurzer Zeit zu reden und zu schweigen, wann und wo sie sollten..... ( )
54.9	*70*	*215*	*53*	*8*	67. Ein guter Schütze trifft selten das Ziel..... ( )
58.8	*68*	*209*	*33*	*50*	68. Es gibt Schmetterlinge von allerlei Farben..... ( )
67.4	*66*	*186*	*35*	*39*	69. Wer zwei Schritte vorwärts geht und einen rückwärts, gelangt schliesslich doch ans Ziel, wenn er nur lange genug geht..... ( )
70.4	*63*	*175*	*34*	*45*	70. Wer ist ein unbrauchbarer Mann? Der nicht befehlen und auch nicht gehorchen kann..... ( )
51.9	*72*	*222*	*51*	*10*	71. Wer seine Schulaufgaben gut vorbereitet, wird gewöhnlich vom Lehrer getadelt ( )
78.4	*54*	*46*	*35*	*40*	72. Um die einheimischen Industrien zu schützen, wird von den Vereinigten Staaten oft ein hoher Zoll auf fremdländische Waren erhoben..... ( )
44.8	*74*	*232*	*43*	*22*	73. Die Butter wird von den Bauern aus Rahm (Sahne) gemacht..... ( )
60.0	*67*	*204*	*33*	*51*	74. Einige Boote und Schiffe werden durch Dampf betrieben, andere durch Segel ( )
52.9	*71*	*220*	*220*	*26*	75. Das kleinste Streichholz brennt heller als die grösste Kerze ..... ( )

GERMAN GRAMMAR COMPLETION TEST

DIRECTIONS: *Each of the English sentences below is followed by a translation which is correct as far as it goes, but which is incomplete. Each translation can be completed correctly by inserting one, two, or three words in each blank* (——). *When you have decided what is necessary to complete the translation, write it on the line at the right of the translation. WRITE LEGIBLY. Do not spend too much time on any one sentence. Do all the easy ones first, then go back to the harder ones if you have time.*

Write carefully and clearly, putting your answers on the lines at the right, as in the examples. Forty-five minutes.

EXAMPLES:

1. My brother is older than I. Mein Bruder ist (——) als ich..... ——— *älter*
2. Your friend has arrived. Ihr Freund (——)..... ——— *ist angekommen*
3. The letter had been delivered. Der Brief (——)..... ——— *war abgeliefert worden*

Per cent correct answers	*Order of difficulty within parts*	*Order of difficulty whole test*	*Index of goodness*	*Order of goodness whole test*	
1	2	3	4	5	
97.4	*1*	*35*	*0*	*267*	1. John and I are brothers. Johann und ich (——) Brüder ———
98.6	*2*	*41*	*0*	*268*	2. The child is ill. (——) ist krank..... ———
95.5	*3*	*55*	*1*	*265*	3. The letter is not from him. Der Brief ist nicht von (——) ———
90.3	*7*	*86*	*11*	*180*	4. Have you the pen? (——) (*Use polite form*) die Feder? ———
85.5	*11*	*111*	*26*	*85*	5. She is a good child. Sie ist ein (——) Kind..... ———
74.5	*23*	*160*	*9*	*196*	6. What was he doing? Was (——) (*Use tun*) er?..... ———
91.6	*5*	*77*	*13*	*164*	7. Why did he do it? Warum hat er es (——) (*Use tun*).. ———

1	2	3	4	5	
90.0	*8*	*89*	*8*	*205*	8. He walked across the bridge. Er (——) (*use imperfect of gehen*) über die Brücke.....
85.5	*12*	*112*	*15*	*154*	9. Who brought the news? Wer hat die Nachricht (——)?
87.8	*9*	*101*	*20*	*123*	10. He is speaking the truth. Er (——) (*Use sprechen*) die Wahrheit.....
91.3	*6*	*81*	*8*	*206*	11. Good people are liked everywhere. (——) Menschen sind überall beliebt.....
82.3	*14*	*123*	*5*	*228*	12. The blackboard. (——) Tafel.....
59.4	*45*	*206*	*18*	*134*	13. The sun was setting. Die Sonne (——) (*Use untergehen*)
85.5	*13*	*113*	*9*	*197*	14. Who took the money? Wer (——) (*Use imperfect*) das Geld?.....
77.1	*20*	*149*	*35*	*41*	15. Who is helping you? Wer (——) Ihnen?.....
77.4	*18*	*147*	*33*	*52*	16. Has he come? (——) gekommen?.....
80.3	*15*	*133*	*29*	*76*	17. A blind man does not see anything. Ein Blinder (——) nichts.....

1	2	3	4	5		
41.0	*67*	*240*	*20*	*124*	18. This note book is mine, that one is yours. Dieses Heft ist meins, jenes ist (——). *(Use polite form)*	
93.2	*4*	*69*	*9*	*198*	19. We heard you singing. Wir hörten Sie (——)	
77.4	*19*	*148*	*28*	*77*	20. If he were here. Wenn er hier (——)	
86.8	*10*	*106*	*3*	*247*	21. What is your name? Wie (——) Sie?	
37.1	*71*	*244*	*23*	*104*	22. The schoolhouse will be built. Das Schulhaus (——)	
66.1	*34*	*191*	*37*	*34*	23. The boy whom we saw. Der Knabe (——) wir sahen	
79.0	*16*	*140*	*15*	*155*	24. You see him now. Sie sehen (——) jetzt	
70.6	*29*	*174*	*24*	*97*	25. She was drinking a cup of milk. Sie (——) eine Tasse Milch	
70.0	*30*	*177*	*13*	*165*	26. This small window. Dieses (——) Fenster	
61.6	*41*	*199*	*30*	*70*	27. We can not do anything with it. Wir können nichts (——) machen	
65.8	*36*	*193*	*17*	*140*	28. He begged me to stay. Er (——) mich zu bleiben	
73.6	*24*	*164*	*25*	*91*	29. I thank you for it. Ich danke Ihnen (——)	

1	2	3	4	5		
76.1	*21*	*151*	*13*	*166*	30. He entered the room. Er (——) (*Use imperfect of treten*) in das Zimmer	
68.0	*31*	*181*	*26*	*86*	31. I shall forget it. Ich (——)	
71.9	*27*	*169*	*5*	*229*	32. The men are gone. Die Männer (——). (*Use perfect of fortgehen*)	
72.6	*25*	*167*	*23*	*105*	33. We study in order to learn. Wir studieren (——)	
13.2	*99*	*274*	*9*	*199*	34. Of all the men I liked Mr. Brown best. Unter all den Männern mag ich Herrn Braun (——)	
58.4	*47*	*210*	*11*	*181*	35. Good books are our best friends. Gute Bücher sind unsere (——) Freunde	
59.7	*44*	*205*	*49*	*12*	36. He repeated the sentences three times. Er (——) (*Use imperfect of wiederholen*) den Satz drei Mal	
61.6	*42*	*200*	*26*	*87*	37. He accepts. Er (——) (*Use annehmen*)	
59.4	*46*	*207*	*45*	*20*	38. Whose knife have you? (——) Messer haben Sie?	
65.5	*37*	*195*	*9*	*200*	39. They slept until eight o'clock. Sie (——) (*Use imperfect*) bis acht Uhr	

1	2	3	4	5		
67.0	*33*	*187*	*14*	*160*	40. The years pass away rapidly. (——) vergehen schnell.	
75.8	*22*	*154*	*20*	*125*	41. The book which you had. Das Buch (——) Sie hatten.	
66.1	*35*	*192*	*25*	*92*	42. Go to the blackboard. Gehen Sie an (——).....	
55.8	*48*	*214*	*57*	*5*	43. Whose house is that? (——) Haus ist das?.....	
44.5	*61*	*234*	*16*	*149*	44. We did remain. Wir (——) (*Use perfect*).....	
42.2	*66*	*239*	*37*	*35*	45. He is said to be very ill. Er (——) sehr krank sein....	
60.6	*43*	*201*	*10*	*188*	46. Those good children. Jene (——) Kinder.....	
71.6	*28*	*171*	*47*	*16*	47. Another book. Ein (——) Buch.....	
63.8	*40*	*198*	*21*	*118*	48. Give me that apple. Geben Sie mir (——) Apfel.....	
46.5	*57*	*228*	*26*	*88*	49. The students whose fathers are here. Die Studenten, (——) Väter hier sind.....	
67.7	*32*	*185*	*15*	*156*	50. Red roses. Rote (——).....	
50.6	*53*	*224*	*64*	*2*	51. The clock was striking ten. Die Uhr (——) zehn.....	

1	2	3	4	5		
54.2	*49*	*216*	*13*	*167*	52. Did you put the book on the table? Haben Sie das Buch auf (——) gelegt?	----------
45.8	*59*	*230*	*43*	*23*	53. The room is ten feet wide. Das Zimmer ist zehn (——) breit	----------
34.5	*76*	*249*	*14*	*161*	54. My garden is large, yours is small. Mein Garten ist gross, (——) (*Use polite form*) ist klein	----------
64.2	*39*	*197*	*30*	*71*	55. My aunt has arrived. Meine Tante (——)	----------
34.5	*77*	*250*	*25*	*93*	56. Do you know where he has been? Wissen Sie, wo er (——)?	----------
47.8	*55*	*226*	*35*	*42*	57. The student whose sister we know. Der Student (——) Schwester wir kennen	----------
41.0	*68*	*241*	*11*	*182*	58. We need those long tables. Wir brauchen diese (——)	----------
32.2	*82*	*258*	*19*	*128*	59. Would you recognize him if you were to see him? (——) ihn erkennen, wenn Sie ihn sähen?	----------
53.6	*50*	*217*	*20*	*126*	60. We would have been able to buy it. Wir hätten es (——)	----------
39.7	*70*	*243*	*9*	*201*	61. The boat had arrived. Das Schiff (——)	----------

1	2	3	4	5	
44.2	*64*	*237*	*22*	*111*	62. Everything that I possess. Alles, (——) ich besitze . . .
53.2	*51*	*218*	*27*	*80*	63. He asked me whether I would go. Er fragte mich, ob ich (——) . . . . .
44.8	*60*	*233*	*22*	*112*	64. The best uncles. Die besten (——) . . . . .
44.5	*62*	*235*	*27*	*81*	65. I like this white paper. Ich habe dieses (——) Papier gern . . . . .
33.5	*80*	*253*	*16*	*150*	66. That is not your bed, but mine. Das ist nicht Ihr Bett, sondern (——) . . . . .
48.7	*54*	*225*	*12*	*173*	67. Do not deceive yourself. (——) (*Use sich täuschen*) . . .
26.5	*87*	*260*	*12*	*174*	68. I have lived in this city since April. Ich (——) in dieser Stadt seit April . . . . .
35.2	*74*	*247*	*22*	*113*	69. There is nobody here. (——) niemand hier . . . . .
64.5	*38*	*196*	*30*	*72*	70. There are many wild animals in these forests. (——) viele wilden Tiere in diesen Wäldern . . . . .
40.3	*69*	*242*	*24*	*98*	71. May we come in? (——) wir eintreten? . . . . .

1	2	3	4	5		
51.6	*51*	*223*	*27*	*82*	72. I was to buy a pair of gloves. Ich (——) (*Use imperfect*) ein Paar Handschuhe kaufen.....	.....
78.7	*17*	*143*	*5*	*230*	73. Come along. (——) (*Use polite form of mitkommen*)...	.....
25.2	*90*	*263*	*16*	*151*	74. I remember my friends. Ich (——) an meine Freunde.	.....
46.5	*58*	*229*	*46*	*17*	75. I should like to see her once more. Ich (——) sie nochmals sehen.....	.....
34.5	*78*	*251*	*1*	*266*	76. He had been seen by the policeman. Er war (——) Polizisten gesehen worden.....	.....
35.8	*73*	*246*	*27*	*83*	77. Had he received my letter, things would be different. (——) meinen Brief erhalten, (——) die Sachen anders stehen.....	.....
27.7	*86*	*259*	*21*	*119*	78. When he had opened the door, the man had disappeared. Als er die Tür (——) (*Use öffnen*), war der Mann verschwunden.....	.....
26.5	*88*	*261*	*26*	*89*	79. The horse has been sold. Das Pferd (——).....	.....
24.5	*91*	*264*	*23*	*106*	80. I was not permitted to do it. Ich habe es nicht (——).	.....
35.2	*75*	*248*	*32*	*59*	81. Who was praised? Wer (——)? (*Use imperfect of loben*)	.....

1	2	3	4	5	
29.7	*85*	*258*	*32*	*60*	82. How many houses have been built? Wie viele Häuser (——)?.....
15.5	*97*	*272*	*33*	*53*	83. He is ashamed of his brother. Er schämt sich (——)..
36.8	*72*	*245*	*18*	*135*	84. You have heard him talk. Sie haben ihn (——).....
26.5	*89*	*262*	*33*	*54*	85. At night we sleep. (——) (*Use an adverb*) schlafen wir.
33.9	*79*	*252*	*31*	*67*	86. The schoolhouse has been built. Das Schulhaus (——) (*Use bauen*).....
32.9	*81*	*254*	*30*	*73*	87. I wish he would go away. Ich wünsche, er (——) (*Use fortgehen*).....
19.7	*92*	*266*	*10*	*189*	88. If you would come early, you would find him at home. Wenn du früh (——), würdest du ihn zu Hause finden
46.8	*56*	*227*	*34*	*46*	89. The children ought to go home. Die Kinder (——) nach Hause gehen.....
30.3	*84*	*257*	*35*	*43*	90. The book can be read. Das Buch kann (——).....
44.5	*63*	*236*	*24*	*99*	91. I am freezing. ——) friert.... .....

1	2	3	4	5		
7.4	*100*	*275*	*16*	*152*	92. I have the names of those who are absent. Ich habe die Namen (——), die abwesend sind.....	----------
19.7	*93*	*267*	*19*	*129*	93. Your work will be done. Ihre Arbeit (——).....	----------
17.4	*95*	*270*	*19*	*130*	94. He succeeded in finding the place. (——) (*Use gelingen in the perfect*) die Stelle zu finden.....	----------
43.5	*65*	*238*	*11*	*183*	95. Before the guests arrived he had gone. Ehe die Gäste kamen (——) (*Use fortgehen*).....	----------
72.6	*26*	*168*	*−4*	*273*	96. Are you talking to me? Sprechen Sie (——)?.....	----------
31.6	*83*	*256*	*32*	*61*	97. What would you like to eat? Was (——) Sie essen?...	----------
17.4	*96*	*271*	*17*	*141*	98. The doors will be opened. Die Türen (——).....	----------
19.7	*94*	*268*	*9*	*202*	99. If these difficulties did not exist. (——) (*Use bestehen*) diese Schwierigkeiten nicht.....	----------
14.5	*98*	*273*	*24*	*100*	100. I should have written. Ich (——).....	----------

The University of the State of New York

233D HIGH SCHOOL EXAMINATION

Thursday, June 18, 1925 — 10.45 A.M. to 12.15 P.M. only

PHYSICS — Part II

READ THESE DIRECTIONS CAREFULLY,
BUT DO NOT TURN OVER THE PAGE UNTIL THE EXAMINER GIVES THE SIGNAL

DIRECTIONS: *Put a plus sign (+) in the parentheses at the right of each statement that is true, and put a zero (0) at the right of each statement that is false. It is better to omit marking statements that you know nothing about. Do not make sheer guesses.*

Do not spend too much time on any one statement. Mark all the easy ones first, then go back to the more difficult ones if you have time.

You will have one hour and thirty minutes to mark all these statements; if you finish before time is called, go back over your work and correct any mistakes that you may have made.

Use a soft lead pencil, a No. 2 if possible. Do not use ink.

Reprinted by Permission.

Per cent correct answers	*Order of difficulty*	*Index of goodness*	*Order of goodness*
1	2	3	4
77.8	*33*	*21*	*88*
50.8	*103*	*14*	*111*
46.5	*114*	*23*	*79*
89.4	*8*	*20*	*91*
52.2	*95*	*34*	*43*
57.5	*82*	*37*	*31*

1. Ten centimeters is approximately four inches.....( )
2. The liter is one-tenth of a cubic meter.....( )
3. The work done by the large piston of a hydraulic press is frequently greater than that done on the small piston.....( )
4. In any enclosed liquid a force applied to any area of the surface is transmitted undiminished to every equal area in the liquid.....( )
5. Atmospheric pressure on the surface of a lake has an effect on the bottom.....( )
6. The pressure due to a depth of 1 meter of water is 100 gr. per sq. cm.....( )

1	2	3	4	
64.2	*66*	*37*	*32*	7. In a vessel filled with water the pressure at any point on the side is the same as at any point on the bottom.....( )
				PROBLEM (*for statements 8 to 11 inclusive*):—**A rectangular box whose base is 20 cm. by 50 cm. is set horizontally and filled to a depth of 10 cm. with water of density 1 gr./cm.^{3}**
76.1	*42*	*41*	*22*	8. The weight of water in the box is 10 kilograms.....( )
51.3	*99*	*7*	*138*	9. The water pressure on the bottom of the box can be found by dividing the weight of water by the area of the bottom.....( )
66.4	*59*	*45*	*12*	10. At a point on the side of the box 10 cm. below the water surface the water pressure is one-half that on the bottom of the box.....( )
86.3	*17*	*27*	*67*	11. To double the pressure on the bottom due to the water pour into the box enough water to make the depth 20 cm.....( )
57.9	*81*	*39*	*28*	12. A heavy body seems to lose more weight in a liquid than a lighter one of the same size, both being completely submerged.....( )
49.1	*109*	*48*	*7*	13. The weight of the liquid displaced by a submerged body is not always determined by the weight of the body.....( )
50.9	*104*	*39*	*29*	14. If the air pressure is increased over the surface of a liquid on which a block is floating, the block will sink deeper in the liquid.....( )
66.8	*58*	*46*	*8*	15. The specific gravity of iron is the same when calculated from measurements in English units as when calculated from measurements in metric units.....( )

1	2	3	4
84.9	*22*	*18*	*97*
51.3	*100*	*14*	*112*
47.8	*110*	*52*	*4*
45.6	*116*	*11*	*132*
85.3	*19*	*34*	*44*
88.0	*10*	*27*	*68*
50.9	*105*	*46*	*9*
76.5	*37*	*14*	*113*
71.2	*50*	*29*	*60*
35.4	*130*	*27*	*69*
68.1	*55*	*11*	*133*
36.3	*126*	*12*	*126*

PROBLEM (*for statements 16 to 19 inclusive*): — **A rectangular block of brass 4 cm. high and 2 cm. by 2.5 cm. in cross-section is suspended, submerged in water, by a thread so that the long edge of the block is vertical. The density of the block is 8.4 gr./cm.**

16. The volume of the block is 20 cm.^{3} ..... ( )
17. The resultant horizontal force on the block is 200 gr ..... ( )
18. The tension in the supporting thread is 20 gr ..... ( )
19. The force of the water on the bottom of the block is necessarily equal to the loss in weight ..... ( )
20. The vacuum above the mercury in a barometer pulls the mercury up into the tube ( )
21. The height to which water can be raised by a lift-pump depends only upon the ease with which the valves operate ..... ( )
22. If the pressure on a given amount of gas is tripled, the density of the gas is also tripled ..... ( )
23. The maximum sum of two forces of 6 lbs. and 4 lbs. is 10 lbs ..... ( )
24. The minimum sum of two forces of 6 lbs. and 4 lbs. is 1 lb ..... ( )
25. It is possible to arrange three forces of magnitudes 3, 5 and 10 lbs., respectively, so that they will produce equilibrium ..... ( )
26. If two objects are dropped at the same time the one which falls for twice the time of the other will travel twice as far ..... ( )
27. Potential energy is measured in the same units as kinetic energy ..... ( )

1	2	3	4	
79.2	*32*	*32*	*47*	28. Potential energy can be changed into kinetic energy ..... ( )
59.3	*77*	*30*	*53*	29. Bodies having equal speeds need not necessarily have equal kinetic energies .... ( )
76.5	*38*	*36*	*36*	30. The rate at which a machine does work depends upon the amount of work to be done ..... ( )
75.7	*44*	*27*	*70*	31. A ten horse-power engine is able to lift 550 lbs. vertically at the rate of 10 feet per second ..... ( )
87.6	*11*	*12*	*127*	32. You should never expect to get from a machine more work than you put into it. ( )
58.8	*78*	*43*	*16*	33. The mechanical advantage of a machine is the reciprocal of the efficiency ..... ( )
96.4	*4*	*16*	*104*	34. Friction in a machine reduces its efficiency ..... ( )
81.9	*28*	*30*	*54*	35. Two moments of force are equal only when the forces themselves are equal ..... ( )
51.8	*97*	*50*	*5*	36. The moment of force due to a given force is independent of the position of the axis of rotation ..... ( )

PROBLEM (*for statements 37 to 44 inclusive*): — **Two girls are on a see-saw, a plank which has been set so that its center of gravity is over the support. One girl, C, weighs 80 lbs. and sits 8 feet from the support. The other girl, D, balances C, although D weighs only 64 lbs.**

86.7	*14*	*27*	*71*	37. Taking the supporting edge as an axis, the moment of force due to C's weight is equal numerically to that due to D's weight ..... ( )
77.8	*34*	*30*	*55*	38. The two girls must exert equal forces on the plank ..... ( )

1	2	3	4
95.0	*5*	*14*	*114*
96.9	*3*	*14*	*115*
50.9	*106*	*41*	*23*
74.7	*47*	*36*	*37*
58.8	*79*	*41*	*24*
67.2	*57*	*46*	*10*
64.2	*67*	*14*	*116*
86.7	*15*	*18*	*98*
53.5	*91*	*9*	*136*
73.4	*48*	*18*	*99*
53.5	*92*	*45*	*13*
65.0	*63*	*27*	*72*

39. C's weight produces a moment of force about the supporting edge of 640 pound feet.....( )
40. D must sit 10 feet from the support.....( )
41. D's weight produces a moment of force of 640 pound feet about a horizontal axis through the point where C sits.....( )
42. The support pushes upward with a force of 72 pounds.....( )
43. If the supporting edge is taken as an axis, the moment of force due to the support is zero.....( )
44. If either girl moves 15 inches farther from the support the other will need to move out the same distance in order to maintain the balance.....( )
45. Drops of water or mud fly off a rapidly rotating wheel in a radial direction.....( )
46. At the curves of a railroad track the outer rail should be higher than the inner rail.....( )
47. The Newton of the "laws of motion" was the father of the Newton of the law of gravitation.....( )
48. The purpose of a barometer is different from that of a thermometer.....( )
49. If two vessels of water have the same amounts of energy their temperatures are the same.....( )
50. An interval of 180° on the Fahrenheit scale is equal to an interval of 100° on the Centigrade scale.....( )

1	2	3	4
46.9	*112*	*9*	*137*
70.0	*52*	*45*	*14*
83.2	*26*	*5*	*141*
54.4	*90*	*21*	*89*
87.1	*13*	*23*	*80*
68.6	*53*	*23*	*81*
54.9	*88*	*34*	*45*
44.2	*118*	*30*	*56*
54.9	*89*	*25*	*75*
84.0	*23*	*16*	*105*
58.8	*80*	*−7*	*144*
77.4	*35*	*46*	*11*
63.2	*68*	*20*	*92*

51. The loss of heat through windows can be decreased by increasing the thickness of the glass.....( )
52. Convection phenomena are found only in liquids and gases.....( )
53. The earth gets heat from the sun only by radiation.....( )
54. If a steel piston fits properly in the steel cylinder of an automobile engine when both are cold, it will not fit properly when both are well heated.....( )
55. A coefficient of linear expansion tells how much a bar of unit length will expand for each degree rise in temperature.....( )
56. Liquids generally expand more than the material of their containers.....( )
57. Absolute zero is at − 273° on the Fahrenheit scale.....( )
58. The pressure of a gas at 60° C is doubled by raising the temperature to 120° C provided there is no change in volume.....( )
59. Change of state from solid to liquid is always accompanied by a decrease in volume.....( )
60. The heat of fusion of ice is about 80 calories per gram.....( )
61. Heat must be added to a liquid in order that it may continue to evaporate at a constant temperature.....( )
62. The temperature at which a liquid boils depends upon the external pressure.....( )
63. Steam can be raised to a temperature above 100° C.....( )

1	2	3	4
65.8	*60*	*16*	*106*
51.3	*101*	*43*	*17*
62.8	*69*	*32*	*48*
46.0	*115*	*7*	*134*
97.3	*2*	*14*	*117*
51.8	*98*	*36*	*38*
29.6	*137*	*50*	*6*
86.3	*18*	*18*	*100*
35.4	*131*	*30*	*57*
36.3	*127*	*14*	*118*
42.9	*120*	*12*	*128*
85.3	*20*	*12*	*129*

64. Heat of vaporization must be added to a liquid before it will change to the vapor state.....( )

65. The change in a Centigrade thermometer always measures the number of calories added to the substance in which the thermometer is placed.....( )

66. Heat energy can not be converted into mechanical energy.....( )

67. The value of the mechanical equivalent of heat has been known for over two hundred years.....( )

68. Sound waves originate at some vibrating body.....( )

69. A sound wave is a compressional wave.....( )

70. The air particles which transmit a sound wave must move with the same speed as that of the wave.....( )

71. The greater the frequency the higher the pitch of a note.....( )

72. The frequency of a sound wave is inversely proportional to the amplitude of the wave.....( )

73. The human ear normally is sensitive to sound waves throughout a range greater than four octaves.....( )

74. Sound waves of the same frequency but of different amplitudes travel at the same speed.....( )

75. The speed of sound in air becomes less as the temperature rises.....( )

1	2	3	4	
15.9	*144*	*14*	*119*	76. A careful observer can detect beats in the sounds received from a source and its echo.....( )
34.5	*134*	*11*	*134*	77. Of two pipes giving the same fundamental the open pipe is shorter than the closed one.....( )
46.9	*113*	*7*	*140*	78. To obtain a note an octave above the fundamental of a violin string put your finger on the middle of the string, and then bow the string.....( )
*79.7**	*31*	*23*	*82*	79. We cannot see an object unless light comes to us from it.....( )
60.6	*74*	*16*	*107*	80. Light from the stars travels much faster than light from terrestrial sources.....( )
61.0	*71*	*29*	*61*	81. A 16-candle-power lamp at a distance of four feet gives the same intensity of illumination as a one-candle-power lamp at one foot.....( )
81.0	*29*	*37*	*33*	82. The angle of incidence is the angle between the incident and the reflected rays..( )
80.1	*30*	*32*	*49*	83. A beam of parallel rays is still a beam of parallel rays after reflection from a plane mirror.....( )
28.8	*138*	*20*	*93*	84. The inverse square law of illumination applies accurately only to point sources..( )
47.3	*111*	*14*	*120*	85. Every spherical mirror cut in any way from a spherical surface of 50 cm. diameter has a radius of curvature of 25 cm.....( )
25.7	*140*	*16*	*108*	86. An object at the center of curvature of a concave mirror will have an image at the principal focus.....( )
35.4	*132*	*29*	*62*	87. The index of refraction of a given material depends upon the angle of incidence.( )

1	2	3	4
39.4	*123*	*25*	*76*
37.6	*124*	*37*	*34*
61.0	*72*	*14*	*121*
45.6	*117*	*14*	*122*
42.9	*121*	*4*	*142*
76.1	*43*	*20*	*94*
75.7	*45*	*25*	*77*
60.6	*75*	*32*	*50*
65.5	*62*	*39*	*30*
36.3	*128*	*29*	*63*
52.7	*93*	*34*	*46*

88. The laws of refraction apply only to refraction at plane surfaces.....( )

89. Total reflection at a surface separating a slow speed medium and a fast speed medium always occurs in the fast speed medium.....( )

90. One surface of a converging lens may be a plane surface.....( )

PROBLEM (*for statements 91 to 92 inclusive*):— **A cube of glass, 3 inches on an edge, with all faces polished, is placed over a black dot on a horizontal sheet of white paper. The index of refraction of the glass is 3/2.**

91. The block appears to be about 2 inches thick.....( )

92. None of the light from the vicinity of the dot is reflected at the upper surface....( )

93. If the thickness of a glass lens is greater at the edges than at the center it is called a diverging lens.....( )

94. Light from the sun focused by a converging lens on a piece of paper forms there an image of the sun.....( )

95. The focal length of a converging lens is the average of object distance and image distance.....( )

96. The lens in a simple camera is a diverging lens.....( )

97. A good compound microscope should give a linear magnification of at least one million.....( )

98. When white light is dispersed by a single prism the yellow rays are unchanged in direction, the red rays are bent one way and the blue the other.....( )

1	2	3	4	
97.8	*1*	*14*	*123*	99. A bar magnet may have only one pole.....( )
68.1	*56*	*12*	*130*	100. Copper may be temporarily magnetized by induction.....( )
62.8	*70*	*29*	*64*	101. A compass needle free to turn about any axis will point toward the North Star..( )
68.6	*54*	*18*	*101*	102. The angle of magnetic declination is usually neglected by land surveyors.....( )
87.6	*12*	*29*	*65*	103. Hard rubber is a good conductor of electricity.....( )
83.6	*25*	*23*	*83*	104. In charging an electroscope by induction be sure to remove the charged body before breaking the ground connection.....( )
52.2	*96*	*23*	*84*	105. The principle of the condenser is the same as that of the electroscope.....( )
44.2	*119*	*45*	*15*	106. Carbon and zinc are most frequently used in cells because they give the highest voltage
72.6	*49*	*43*	*18*	107. Some voltaic cells are used to supply alternating current.....( )
57.5	*83*	*41*	*25*	108. The effect of cell polarization is most noticeable when the circuit is first closed..( )
74.7	*47*	*36*	*39*	109. If one of the plates of a voltaic cell is zinc, the material of the other has no effect on the electromotive force of the cell.....( )
65.8	*61*	*59*	*1*	110. In a simple circuit of voltaic cell and external resistance the current can not flow when the circuit is "closed".....( )
				PROBLEM (*for statements 111 to 116 inclusive*):— **A lead storage cell is to be charged from the direct current mains. A controlling rheostat is used in series with the cell.**
55.3	*86*	*16*	*109*	111. The positive plate of the storage cell should be connected to the positive terminal of the mains.....( )

1	2	3	4	
65.0	*64*	*41*	*26*	112. The current entering the cell while it is charging is necessarily smaller than that leaving the cell during the same time.....( )
51.3	*102*	*14*	*124*	113. The storage cell has a back electromotive force during the charging.....( )
90.2	*7*	*12*	*131*	114. A chemical change is taking place during the charging.....( )
61.0	*73*	*59*	*2*	115. The current in the cell must be in the same direction whether the cell is being charged or discharged.....( )
36.3	*129*	*23*	*85*	116. It makes no difference to which side of the cell the rheostat is connected.....( )
77.0	*36*	*32*	*51*	117. The magnetic field about a conductor exists only when the current is changing..( )
85.3	*21*	*23*	*86*	118. If an electric current is reversed in direction the accompanying magnetic field is also reversed in direction.....( )
84.0	*24*	*36*	*40*	119. The best kind of iron for the core of an electromagnet is that used in permanent magnets.....( )
76.5	*39*	*43*	*19*	120. An electromagnet must have primary and secondary windings.....( )
26.1	*139*	*41*	*27*	121. The electromagnet of a relay in a telegraph line is excited by the "local battery".( )
70.8	*51*	*37*	*35*	122. Incandescent lamps in house circuits are usually connected in parallel.....( )
50.4	*107*	*32*	*52*	123. If a copper wire and an iron wire of the same length have the same resistance the diameter of the copper wire must be larger than that of the iron wire.....( )
76.5	*40*	*54*	*3*	124. The resistance of an ordinary lamp filament depends upon the direction of the current in the filament.....( )

1	2	3	4	
86.7	*16*	*20*	*95*	125. Electric current is measured in amperes ..... ( )
76.5	*41*	*36*	*41*	126. When we purchase electrical energy we pay for a certain number of amperes .... ( )
50.4	*108*	*20*	*96*	127. The amount of current consumed by a 25 watt lamp is not the same on every circuit ..... ( )
88.4	*9*	*16*	*110*	128. A lamp of 240 ohms resistance takes one-half ampere on a 120 volt circuit..... ( )
34.9	*133*	*18*	*102*	129. We cannot get power from a generator unless the current entering the generator from the line is smaller than that entering the line from the generator..... ( )
52.7	*94*	*30*	*58*	130. A customer of a commercial power company cannot use one kilowatt-hour in less than one hour..... ( )
57.5	*84*	*43*	*20*	131. If two heating coils, one of twice the resistance of the other, be connected in series to the 120 volt main, the greater amount of heat will be developed in the coil of smaller resistance..... ( )
82.3	*27*	*27*	*73*	132. The dial of a voltmeter is frequently calibrated to read ohms..... ( )
40.7	*122*	*36*	*42*	133. The ammeter in an electrical circuit is not analogous to the pressure gauge in a water system ..... ( )
65.0	*65*	*11*	*135*	134. If you reverse the armature connections of a simple D. C. shunt motor the direction of rotation will be reversed..... ( )
57.5	*85*	*43*	*21*	135. An induced current is necessarily in the same direction as the primary current.. ( )
16.4	*143*	*25*	*78*	136. It is possible to move a coil of wire in a uniform magnetic field without inducing an electromotive force ..... ( )

1	2	3	4	
60.2	*76*	*30*	*59*	137. The average speed of the molecules of a gas is much less than that of those in the liquid state at the same temperature.....( )
91.6	*6*	*21*	*90*	138. Newton's laws of motion are no longer considered to be even approximately correct.....( )
31.9	*135*	*29*	*66*	139. A body is said to be radiating when and only when it is emitting electrons.....( )
25.2	*141*	*18*	*103*	140. It is inconsistent with the principle of conservation of energy to say that heat energy may become unavailable.....( )
31.9	*136*	*27*	*74*	141. Electrons flow through a perfect condenser.....( )
55.3	*87*	*14*	*125*	142. A copper atom differs from a copper ion in the number of its electrons.....( )
24.3	*142*	*0*	*143*	143. The period of an oscillating electrical circuit can be decreased by using a smaller capacitance.....( )
37.6	*125*	*23*	*87*	144. An electron moving south in an east-to-west magnetic field will be deflected to the east.....( )

Size of vocabulary samplings. — A fundamental factor in the validity of any examination is the size of the sampling of materials and tasks which it involves. A simple illustration of this principle may be taken from the field of spelling. Every one knows that a spelling test of only ten words will not give an adequate measure of an individual's English spelling ability. A 50-word spelling test might give a fairly reliable indication of the average spelling ability of a whole class, but there are few scholars who would base their judgment of an individual's spelling ability on the results of fifty words. Other things being equal, the larger the number of words in a spelling test, the more accurate and reliable is the resulting indication of an individual's spelling ability. The phrase "other things being equal" here covers a multitude of sins because the character of the sampling is quite as important as the size; but for the moment let us keep in mind only the matter of size of sampling. If the new-type form of examination permits us to use a larger sampling of modern language materials than the old-type, other things remaining equal, the new-type is a more valid form of examination.

Vocabulary sampling of ninety-minute new-type test about twice as large as that of three-hour old-type examination. — A vocabulary analysis of the old-type Regents and College Entrance Examination Board French examinations of June, 1924, and June, 1925, shows that the numbers of different French words involved in old-type three-hour examinations, exclusive of the English-to-French translation and of the free composition questions, range from 206 to 270 root words. The new-type 90-minute Regents examination in French of June, 1925, includes 459 different French words, and the new-type 90-minute junior high school French examination of June, 1925, involves 529 different root words. Inflected forms of the same root words were not counted as additional words; that is, one verb might appear in two or three tenses and in two or three numbers, but these were all counted as one root word.

That the new-type 90-minute examinations permit a sampling of root words roughly twice as large as the sampling permitted by the old-type three-hour examinations is a fact of tremendous significance. This fact alone is sufficient to account for many of the advantages of the new-type, in terms of reliability and validity, pointed out in earlier sections of this report. It should be noted, however, that this comparison relates only to size of vocabulary sampling and, moreover, that it does not take into account two important parts of the old-type examinations, namely, the English-to-French translation question and the free composition question. These two questions probably would add between 50 and 75 different root words to the numbers reported above for old-type three-hour examinations. But since these two questions are generally allowed about a third of the total credit of an old-type three-hour examination and

therefore presumably take up about one of the three hours of time, we may leave the numbers as reported above and bear in mind that we are comparing a two-hour old-type with a 90-minute new-type examination.

That this comparison is limited to size of vocabulary only should not lead us to underestimate its importance. It may be objected, for example, that although the old-type includes a much smaller sampling of different root words, it makes up for this deficiency by the character of the words used and by the richness of the grammatical and the idiomatic contexts in which these words appear. The answer to this objection is that the display of grammatical and idiomatic knowledge by a student depends quite as much upon the specific vocabulary presented to him as upon his knowledge of the principles of grammar involved. Many idioms, for example, are inseparable from certain specific words, so that there is little room for doubt that any limitation of vocabulary sampling is necessarily a measurable limitation of grammar and idiom sampling.

Character of vocabulary sampling. — The objection mentioned in the preceding paragraph, that the character of the vocabulary sampling in the old-type might be such as to make up for its small size, would be very serious indeed if supported by evidence. It is quite obvious that merely increasing the number of very easy words which all students know, or of very difficult words, which very few students know, does not increase the effective sampling or the real validity of an examination. The existence of Professor Henmon's French Word Book [1] and of the Wood French word list [2] makes possible an objective study of the character of the vocabulary samplings in the old- and new-type French examinations with which we are now concerned. Table 37 shows ordinary and cumulative distributions, according to frequency of occurrence, of the 3900 words in the Henmon French Word Book and of the 2683 words in the Wood French word list.

Henmon and Wood French Word Frequency Distributions. — According to the Henmon French Word Book there are 153 different root words which occur between 200 and 27,000 times in 400,000 words of running discourse; and according to the Wood French word list there are 134 root words which are common to 16 widely used French textbooks. These words are the essential words of the language which all students necessarily learn in the first year of modern language work, if they learn anything at all about the modern language. As far as vocabulary sampling in a test is concerned, the inclusion of a large number of these easy words would be wasteful and would not add to the effectiveness of the sampling. Certainly not more than 5 or 10% of the different words in a test should

[1] Henmon, V. A. C., *A French Word Book*, University of Wisconsin Bulletin No. 3, Sept. 1924. This study has been extended by Ward (C. F.) in his book entitled *Minimum French Vocabulary Test Book*.

[2] Wood, Ben D., *A Comparative Study of the Vocabularies of Sixteen French Textbooks*, The Modern Language Journal, February, 1927.

come from this group. For third- and fourth-year students, words from this group probably have no measurement value *as far as vocabulary alone* is concerned. Similarly, words which occur fewer than five times in 400,000 words of running discourse, or words which are common to fewer

TABLE 37

The Henmon and Wood French Word Lists

DISTRIBUTIONS OF 3900 ROOT WORDS OCCURRING FIVE TIMES OR OFTENER IN 400,000 WORDS OF RUNNING DISCOURSE ACCORDING TO FREQUENCY OF OCCURRENCE, AND OF 2683 WORDS COMMON TO FOUR OR MORE OF SIXTEEN TEXTBOOKS ACCORDING TO NUMBER OF BOOKS TO WHICH THEY ARE COMMON. THE TABLE IS TO BE READ AS FOLLOWS: FIRST LINE — 59 ROOT WORDS OCCURRED BETWEEN 500 AND 27,000 TIMES IN 400,000 WORDS OF RUNNING DISCOURSE; AND 134 WORDS WERE FOUND COMMON TO SIXTEEN WIDELY USED GRAMMARS AND COMPOSITION BOOKS. SECOND LINE — 94 WORDS OCCURRED 200–499 TIMES, AND 153 WORDS OCCURRED 200 TIMES OR OFTENER IN 400,000 WORDS OF RUNNING DISCOURSE; 110 WORDS WERE COMMON TO FIFTEEN, AND 244 TO FIFTEEN OR SIXTEEN TEXTBOOKS. SEVENTH LINE — 1035 WORDS OCCURRED 10–19 TIMES; AND 2533 WORDS OCCURRED 10 TIMES OR OFTENER IN 400,000, ETC.; 389 WORDS WERE COMMON TO FIVE TEXTS, AND 2208 WERE FOUND COMMON TO FIVE OR MORE OF SIXTEEN TEXTBOOKS.

THE CORRELATION BETWEEN HENMON AND WOOD FREQUENCIES IS 0.70.

HENMON FRENCH WORD BOOK				WOOD FRENCH WORD LIST		
Henmon Frequencies		f_H	Cumulative f_H	Wood Frequencies	f_W	Cumulative f_W
(1)	500–27,000	59	59	16	134	134
(2)	200–499	94	153	15	110	244
(3)	100–199	169	322	14	142	386
(4)	60–99	210	532	13–12	240	626
(5)	40–59	280	812	11–9	463	1089
(6)	20–39	686	1498	8–6	730	1819
(7)	10–19	1035	2533	5	389	2208
(8)	5–9	1367	3900	4	475	2683
$r_{HW} = 0.70$						

than four of sixteen widely used textbooks, almost certainly have very little measurement value for first- and second-year students. We should therefore expect the second-year old-type examinations to contain larger proportions of the very easy words than the third- and fourth-year old-type examinations, and the latter to contain larger proportions of the more difficult words.

Table 38 shows the per cents of words from each of the Henmon and Wood frequency groups in the old-type French examinations, and Table 39 shows similar per cents for the new-type Regents and junior high school French examinations of June, 1925.

New-type superior to old-type vocabulary sampling according to Henmon and Wood French word lists. — Table 38 shows that the character of the vocabulary samplings in the old-type examinations is in every way inferior to that of the new-type samplings, if the Henmon and Wood French word lists have any validity at all. There is practically no difference between the difficulties of the vocabulary samplings of the second-, third- and fourth-year old-type examinations. There is practically as large a proportion of very easy words in the fourth-year examination as in the

TABLE 38

DISTRIBUTIONS ACCORDING TO HENMON AND WOOD FREQUENCIES OF ROOT WORDS IN OLD-TYPE REGENTS AND COLLEGE ENTRANCE EXAMINATION BOARD EXAMINATIONS IN SECOND-, THIRD- AND FOURTH-YEAR FRENCH OF 1924 AND 1925. THESE ARE ALL THREE-HOUR EXAMINATIONS, EXCEPT THE 1925 REGENTS, WHICH ARE 90-MINUTE EXAMINATIONS. THE FREE COMPOSITION QUESTIONS IN ALL THESE EXAMINATIONS ARE, OF COURSE, NOT ACCOUNTED FOR IN THIS TABLE; THE ENGLISH-TO-FRENCH TRANSLATION QUESTIONS ARE ACCOUNTED FOR ONLY IN THE CASE OF THE REGENTS 1925 EXAMINATIONS. (REGENTS 1924 FOURTH-YEAR EXAMINATION WAS NOT AVAILABLE FOR THIS STUDY.)

Regents Old-Type Examinations

HENMON FREQUENCIES	1924 3-HOUR EXAMINATIONS						1925 90-MINUTE EXAMINATIONS					
	2nd Year		3rd Year		4th Year		2nd Year		3rd Year		4th Year	
	N	%	N	%	N	%	N	%	N	%	N	%
500+	52	19.2	46	22.2			40	23.5	44	26.5	42	27.4
100–499	81	29.9	55	26.4			56	33.0	48	28.8	40	26.1
40–99	52	19.1	31	14.8			26	15.3	25	15.0	28	18.4
20–39	31	11.5	22	10.6			19	11.2	17	10.3	17	11.1
5–19	30	11.1	30	14.5			22	12.9	20	12.1	14	9.2
0–4	25	9.2	24	11.5			7	4.1	12	7.3	12	7.8
Totals	271		208				170		166		153	
WOOD FREQUENCIES												
16	68	25.1	52	25.0			54	31.8	50	30.2	51	33.3
15	40	14.8	28	13.5			30	17.6	24	14.6	22	14.4
14	26	9.6	17	8.2			19	11.2	19	11.5	18	11.7
13–12	43	15.8	23	11.2			20	11.8	18	10.8	15	9.8
11–9	29	10.7	18	8.6			16	9.4	13	7.8	13	8.5
8–6	19	7.0	17	8.1			13	7.6	14	8.4	7	4.6
5–4	12	4.4	17	8.2			6	3.6	4	2.4	7	4.6
3–0	34	12.6	36	17.3			12	7.0	22	13.3	20	13.1
Totals	271		208				170		166		153	

College Entrance Examination Board Examinations

Henmon Frequencies	1924 3-Hour Examinations						1925 3-Hour Examinations					
500+	49	23.8	50	21.3	47	19.5	52	21.9	48	18.9	46	17.0
100–499	51	24.7	57	24.3	68	28.2	75	31.7	72	28.6	80	29.6
40–99	37	17.9	45	19.1	47	19.5	51	21.5	51	20.2	36	13.4
20–39	30	14.6	30	12.8	23	9.5	20	8.4	24	9.4	38	14.1
5–19	20	9.7	32	13.6	36	15.0	23	9.7	29	11.4	41	15.2
0–4	19	9.3	21	8.9	20	8.3	16	6.8	29	11.5	29	10.7
Totals	206		235		241		237		253		270	
Wood Frequencies												
16	62	30.1	61	26.0	55	22.8	72	30.3	64	25.3	58	21.5
15	24	11.7	33	14.0	32	13.2	31	13.0	44	17.4	37	13.7
14	21	10.2	20	8.5	20	8.3	32	13.5	28	11.1	23	8.5
13–12	24	11.6	30	12.7	30	12.5	28	11.8	28	11.0	31	11.5
11–9	18	8.8	23	9.9	33	13.7	26	11.1	19	7.5	31	11.5
8–6	21	10.2	20	8.5	20	8.3	15	6.4	24	9.5	23	8.5
5–4	9	4.3	14	6.0	18	7.5	9	3.8	10	4.0	23	8.5
3–0	27	13.1	34	14.4	33	13.7	24	10.1	36	14.2	44	16.3
Totals	206		235		241		237		253		270	

TABLE 39

Distributions according to Henmon and Wood frequencies of root words in the new-type junior high school and Regents examinations in French of June, 1925. These two examinations are identical in general form, but Part II, reading-comprehension, is made up of multiple choice questions in the junior high school and of true-false questions in the Regents examination.

New-Type Examinations in French of June, 1925

Henmon Frequencies	Junior High School: 90 Minutes							Regents: 90 Minutes						
	Pt. I	Part II		Part III		Whole Test		Pt. I	Part II		Part III		Whole Test	
	N / %	N	%	N	%	N	%	N / %	N	%	N	%	N	%
500+	3	53	12.0	28	37.5	54	10.2	1	51	14.9	30	44.1	55	12.0
100–499	19	99	22.4	21	28.0	115	21.7	15	96	27.9	17	24.0	117	25.5
40–99	19	109	24.6	12	15.9	124	23.4	9	66	19.1	7	10.4	80	17.4
20–39	17	62	14.1	9	12.0	79	14.9	17	63	18.3	7	10.3	81	17.7
5–19	27	73	16.5	1	1.3	95	18.0	38	47	13.7	6	8.8	86	18.7
0–4	15	46	10.4	4	5.3	62	11.8	20	21	6.1	1	1.5	40	8.7
Totals . .	100	442		75		529		100	344		68		459	
Wood Frequencies														
16	11	91	20.7	30	40.0	98	18.5	5	76	22.1	30	44.1	87	18.9
15	9	57	12.9	11	14.7	62	11.7	8	46	13.3	8	11.8	51	11.1
14	10	49	11.1	10	13.3	59	11.2	6	45	13.1	7	10.3	53	11.6
13–12	18	82	18.5	9	12.0	94	17.7	4	57	16.6	12	17.6	65	14.1
11–9	2	71	16.0	5	6.7	74	14.0	4	53	15.3	7	10.2	60	13.1
8–6	29	43	9.7	3	4.0	79	13.1	34	22	6.4	2	3.0	58	12.6
5–4	21	16	3.6			35	6.7	39	9	2.7	1	1.5	48	10.5
3–0		33	7.5	7	9.3	38	7.2		36	10.5	1	1.5	37	8.1
Totals . .	100	442		75		529		100	344		68		459	

second-year examination; and the proportions of difficult words in the second-year examinations are almost as great as in the fourth-year examinations. Some of the opponents of the objective tests are particularly suspicious of the value of the new-type vocabulary test. The statement is frequently made that they "do not care about *mere* vocabulary." Yet 10% of the words in the old-type second-year examinations occurred fewer than five times in 400,000 words of running discourse and 15% of them are common to fewer than four of sixteen widely used textbooks; in other words, the makers of old-type tests go into the fifth thousand of French words in their second-year examinations. This is indeed "not caring about vocabulary" with a vengeance.

The facts of Tables 38 and 39 are summarized graphically in Chart 44. The indications for these old-type examinations may be accepted as thoroughly representative of old-type French examinations generally, because they are based on analyses of eight full three-hour examinations, excluding English-to-French and free composition questions, and on three ninety-minute old-type examinations, excluding the free composition questions. The *proportions* of words in these examinations from the several Henmon and Wood frequency groups would probably not be observably changed by the inclusion of the English-to-French and free composition questions. The proportions for the three ninety-minute Regents examinations, in which the English-to-French questions are included in the vocabulary analysis, are not markedly different from those for the other old-type examinations.

It is apparent from Chart 44 that all of the old-type examinations include almost equal proportions of words in the highest Henmon frequency groups, that is, of words that occur one hundred times or oftener in 400,000 words of running discourse. Each of the old-type examinations contain 42 or more of the 59 words that occur 500 times or oftener in 400,000 words of discourse. Such words can have little if any measurement value, *as vocabulary sampling*, except possibly in the first- and second-year classes. That they represent from a fifth to a fourth of the total number of different root words in the fourth- and third-year old-type examinations may therefore be taken to mean that the *effective vocabulary sampling* of the old-type examinations for these year-classes is only four-fifths or three-fourths as large as the total numbers of different root words in them, that is, an average of less than 200 rather than an average of 240 words. It may be argued, perhaps with considerable reason, that these fundamental and indispensable words are essential for the testing of other elements such as grammar, idioms, syntax, etc., in the higher years; but this argument cannot even remotely be used in justification of the nearly equally large proportions of very difficult words (from the opposite end of the frequency scale) in the second-, third- and fourth-year examinations.

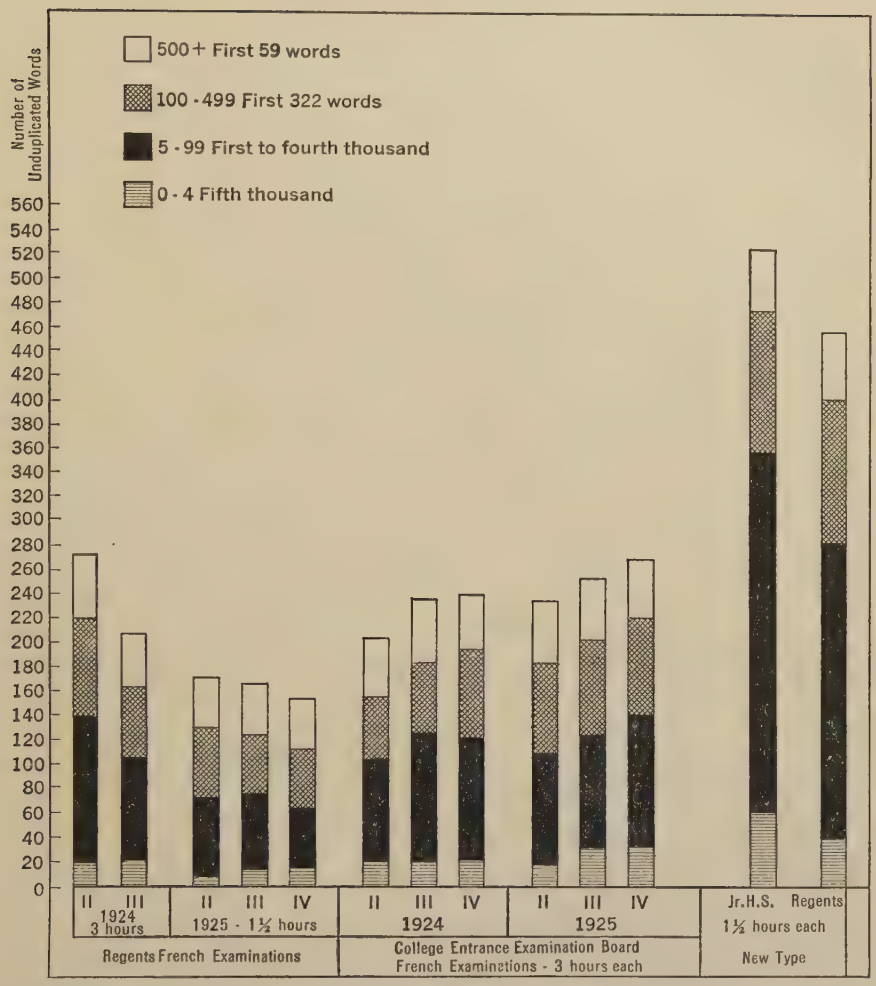


CHART 44. — Numbers of different root words in old- and in new-type tests distributed according to Henmon frequency groups.

Chart 44 shows that, on the whole, the second-year examinations contain as large proportions of words not in the Henmon Word Book as the third- and fourth-year examinations do. It is difficult to find any plausible excuse for the inclusion of almost as many fifth-thousand words in the French II examinations as in the French IV examinations. There are certainly not many grammatical or idiomatic forms that second-year students ought to know which depend on the use of words that will not be encountered five times in 400,000 words of general reading matter. It seems more likely that the inclusion of the rare words in the second-year examinations, and also the inclusion of the very easy words in the

fourth-year examinations, is due to the chance factors inherent in the subjective and unscientific method of constructing old-type examinations.

Not infrequently one hears modern language teachers say that in their opinion a certain old-type examination is easier than the one designed for the next lower class, or harder than the one for the next higher class. And it is a matter of record that considerable numbers of students sometimes pass a third- or fourth-year examination on the same day that they fail a second- or third-year examination.[1] Chart 44 affords objective evidence that the old-type examinations for second-, third- and fourth-year classes are at least very nearly equal, if not in fact slightly inverted, in difficulty as far as vocabulary is concerned. That such differences as do exist are slight, and do not always tend to make the higher examinations harder than the lower, is clearly indicated in Chart 45.

Old-type second- and fourth-year examinations nearly equal as to vocabulary difficulty and probably also as to grammar and syntax. — The importance of this indication is that the second-, third- and fourth-year examinations are very probably as nearly equal in respect to grammar, idiom and syntax difficulties as in respect to vocabulary difficulties. This inference seems justified for at least two reasons. In the first place, the method of selecting the grammar, idiom and syntax elements for the old-type examinations is just as subjective as the method of selecting the vocabulary content, and there is no reason for supposing that the modern language scholars who make up the old-type examinations would make more errors in judging the difficulty and measurement value of vocabulary elements than of grammar and idiom elements. In the second place, even if the actual rules of grammar and syntax and idioms in the elementary examinations are easier than in the advanced, their greater easiness might be effectually masked by the strange and unknown words which are their vehicles. It is certainly conceivable that a student might miss a perfectly simple point of grammar not because he is ignorant of it, but because it occurs in a context of words which will not be encountered five times in 400,000 words of French discourse, and which are not common to more than three of sixteen widely used textbooks. With such nearly equivalent examinations being used for second-, third- and fourth-year classes, we can no longer be surprised at the large overlapping of classes set forth earlier in this report. Students of only second-year ability in French who happen to be in a French IV class have almost as good a chance of passing the fourth-year examination as of passing the second-year examination. In view of these facts it is idle to suggest that some of the weaknesses of the old-type examinations set forth in these pages, are in reality due to

[1] A member of the Advisory Board of the Regents Examinations told the writer of this Report that a few years ago the Board interchanged the names of the third- and fourth-year papers submitted by the Committee for one of the modern languages.

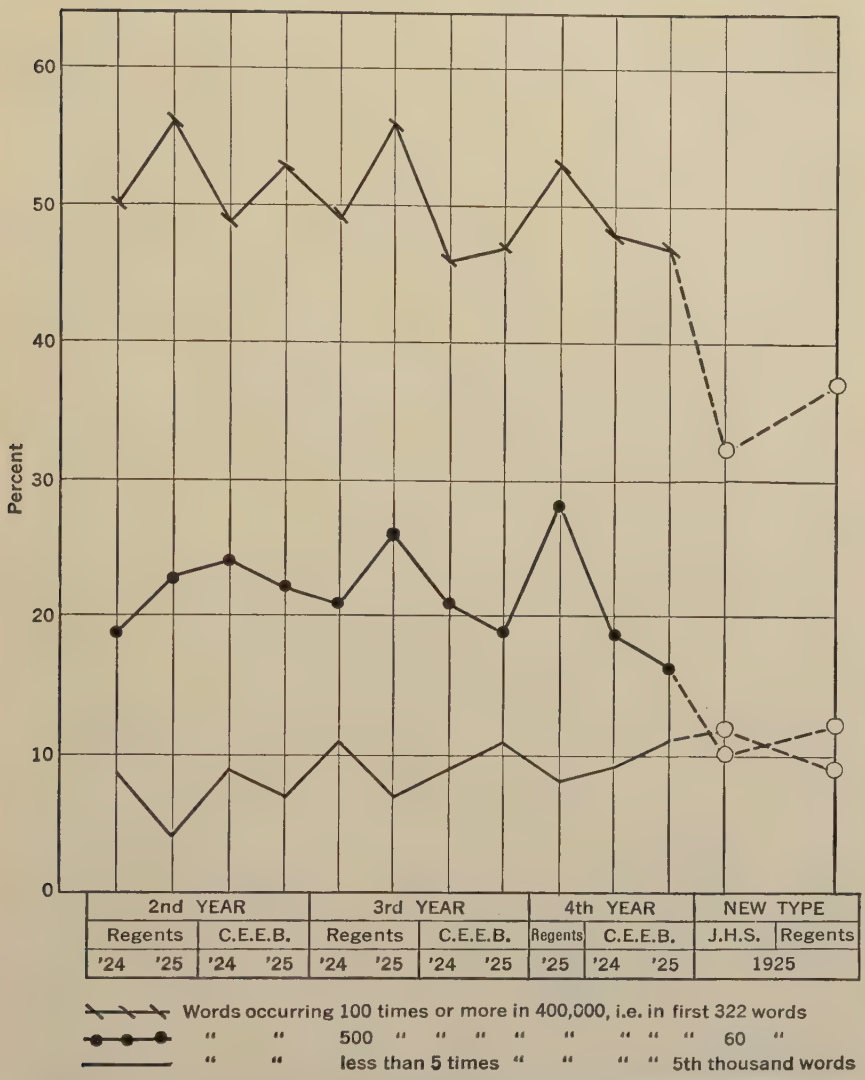


CHART 45. — Proportions of root words in old- and in new-type French examinations of indicated Henmon frequencies, showing that there are only small differences between the vocabulary difficulties of 2nd- and 4th-year old-type examinations, and that all old-type French examinations here represented contain large proportions of words that have little measurement value, so far as vocabulary is concerned. The 1925 Regents old-type and both new-type tests are 90-minute examinations; all others are 3-hour examinations. This chart is made from the data of Tables 38 and 39, *q.v.*

the invalidity of the new-type test which has been used as a sort of check on the other type.

Proportions of high- and low-frequency words in Regents and College Entrance Examination Board old-type examinations and in Regents and junior high school new-type tests. — In Chart 45 the old-type examinations are arranged in ascending order of year-classes from left to right, with the new-type French examinations at the extreme right. The dash-lines connecting the points of the new-type with the curves of the old-type are put in merely to identify the corresponding frequency groups; they should not be thought of as continuations of the curves. Both the general contours and positions, and the jagged nature of the curves, show that chance had considerable influence on the character of the vocabulary samplings of the old-type examinations.

The lowest of the three curves shows the proportion in each of the examinations of words which occurred less than five times in 400,000 words of discourse. These are words from the fifth thousand, or higher thousands, of French words according to frequency of use in general French reading matter. Such rare words certainly have a place in fourth-year examinations, but they can have very little measurement value, if any, in first- and second-year examinations, and they are of doubtful value in a third-year high school examination. We should therefore expect that there would be very few of these words in second-year examinations and a moderate number in fourth-year examinations; in other words, that the lowest curve would ascend from near the base line at the left to about 10% at the right. As a matter of fact, however, the curve, in spite of irregularities, is practically flat in general trend. About 11% of the words in two of the four old-type third-year examinations represented in the chart, are in this category; in only one of the three fourth-year examinations is the proportion of such words as great as 11%, the proportions in the other two being 8 and 9 per cent. Two of the four second-year examinations have proportions of 9% of such words. In other words, we are asking second-year students to know as many fifth-thousand words as we ask fourth-year students to know, and are also asking second-year students to make their knowledge of grammar, idioms, syntax, etc., shine through the mists of these rare words, or remain invisible, unhonored and unsung, but not unwept.

The middle curve shows the proportion of words in the old- and the new-type tests from the group of 59 words which occur 500 times or oftener in 400,000 words of running discourse. Words from this group might logically be expected in first- and second-year old-type examinations but they probably have little, if any, measurement value for students above the second-year level. We should therefore expect the middle curve to slant downward to the right since the old-type examinations are arranged

in order of year-classes from left to right. The examination which shows the highest proportion of such easy words, however, is a fourth-year examination and the second-highest proportion is shown by a third-year examination. One of the second-year examinations is among the four old-type papers that show the lowest proportions of such easy words.

The topmost curve in Chart 45 shows for each of the examinations the proportions of words that occur 100 times or oftener in 400,000 words of running discourse. According to the Henmon Word Book there are only 322 different French root words that occur 100 times or oftener. We would expect to find larger proportions of the first 300 words in the second-year examinations than in the fourth-year examinations, that is, if the examinations were constructed in the best possible manner we would expect the topmost curve in Chart 45 to slant downward to the right. But here again chance seems to have had as much influence as in the choice of the very difficult words. Forty-nine per cent of the different root words in the 1924 College Board second-year French examination are from the first 322 words according to the Henmon list, and corresponding per cents for the two fourth-year College Board papers are 47 and 48. The examination which has the lowest proportion of the first 322 words is the College Board third-year paper of 1924.

The proportions of easy and difficult words of new-type tests were also determined on the basis of objective evidence. The proportion of fifth-thousand words for the Regents new-type examination is 9% and for the junior high school 12%. The larger proportion of rare words in the junior high school test is due to the fact that the alternative answers in the multiple choice form of reading test were purposely allowed to include, in a few cases, unusual words. Twelve per cent of the words in the new-type tests are from the group of 59 words which occur 500 times or oftener in 400,000 words of running discourse, and 32% and 37% of the new-type test words are from the first 322 words according to the Henmon French Word Book. In an examination which is designed to cover the whole range of achievement in high school modern language work the proportions for the new-type tests seem entirely reasonable.

Chart 46 is parallel to Chart 45 in general form, and shows the proportions for each of the old- and new-type examinations of words which are common to fewer than four of sixteen textbooks, words which are common to 14, 15 and 16 textbooks, and words which are common to 16 textbooks. The lowest curve in Chart 46 shows that there is practically no difference between third- and fourth-year old-type examinations and only negligible differences between these and the second-year old-type examinations with respect to proportions of words common to three, or fewer, of sixteen textbooks. Two of the old-type examinations show proportions of such words as high as 17%; one of these is a third-year and the other a fourth-

year paper. Two of the four second-year old-type examinations have proportions of very rare words of 13%; one third-year and two of the fourth-year old-type examinations show proportions of 13%. It is interesting to note the differences between examinations for the same year-class in different years. In the second-year Regents examination of 1924, 13% of the words were common to fewer than four of sixteen textbooks and in the 1925 second-year examination only 7% of the words fell in this category. These facts go far towards accounting for the variations in standards of the Regents and College Board examinations and also in accounting for the tremendous overlapping of classes. More than one in six of the different root words in the 1924 Regents French examination for the third-year are words which are common to fewer than four of sixteen widely used textbooks; in two of the three fourth-year examinations represented in Chart 46 only about one in eight of the different words are in this category. We can no longer wonder at the fact that students sometimes pass a fourth-year examination on the same day that they fail on a third-year examination. That the examinations used by the highest and most universally recognized standardizing agencies in the country should show such chance variations as these is a notable commentary on the extent to which American educators have underestimated both the difficulty and the importance of constructing sound measuring devices.

The middle curve of Chart 46 shows for each of the examinations the proportions of words common to sixteen textbooks, that is, of the first 134 words according to the Wood French word list. The examination which shows the highest proportion of these very easy words is the fourth-year Regents examination of June, 1925. The examinations which contain the lowest proportions of such words are the fourth-year papers of the College Board for 1924 and 1925. The Regents 1924 second-year paper shows a proportion of these words as low as 25%, which is the size of the proportion of the third-year papers of the Regents of 1924 and of the College Board of 1925.

The topmost curve in Chart 46 shows the proportions for each examination of words common to 14, 15 and 16 textbooks. According to the Wood French word list there are less than 400 words in this category. Again we would expect that the fourth-year examinations would contain smaller proportions of these words than second-year examinations contain. The curve does show a slight tendency to slant downward to the right but it also shows notable variations. Fifty-nine per cent of the words in the Regents fourth-year paper of June, 1925, and 52% of the words in the 1925 College Board fourth-year paper, fall in this category. In other words, the 1925 fourth-year College Board French examination contains a larger proportion of the first 400 words, according to the Wood

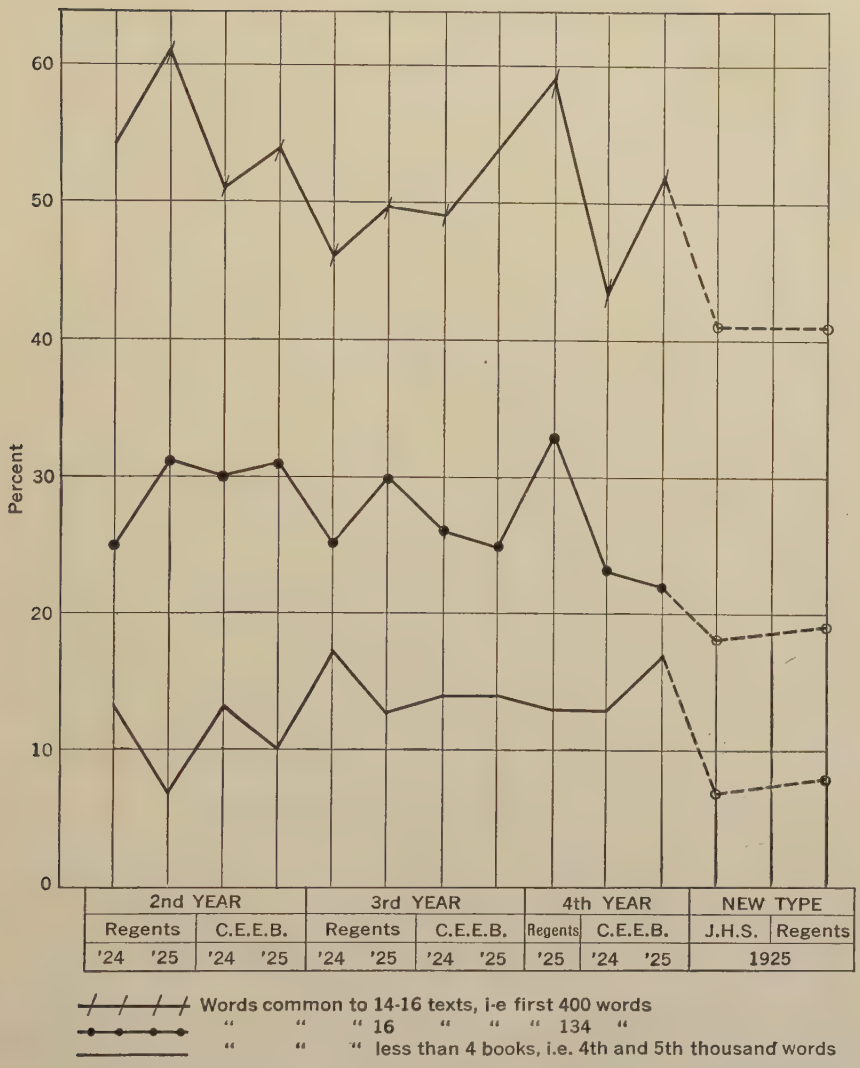


CHART 46. — Proportions of root words in old- and in new-type French examinations common to indicate numbers of textbooks. This chart is similar to Chart 45, *q.v.*

French word list, than did the College Board 1924 second-year French examination. It would be superfluous to comment further on these variations. The proportions of words from these three categories in the new-type French examinations are such as would be expected in a test designed to cover the whole range of achievement in modern language work. Only about 8% of the different root words in the new-type papers are common to fewer than four of sixteen widely used textbooks. In

other words, all of the students for whom the new-type tests were designed, may reasonably be expected to have been exposed to at least 90% of the root words in them. It would certainly be unfair to test students below the fourth-year level with examinations containing a larger proportion of words not found in three out of four of the textbooks used in French classes. Less than a fifth of the words in the new-type tests are from the first 134 words, that is, from those common to 16 textbooks; and about two-fifths of them are from the first 400 words, according to the Wood French word list. A proportion of 20% of these words in an examination designed for the whole range of high school modern language achievement seems much more reasonable than a proportion of 33% of such words in an examination designed solely for fourth-year students, such as the 1925 Regents examination in French.

Empirical difficulties of individual questions in new-type tests. — The empirical difficulties of each individual question in the new-type examinations used in this experiment have been ascertained on the basis of the per cents of correct responses from random groups of students ranging in numbers from 200 to 1000. The first column of figures at the left of the questions in the new-type examinations reproduced above on pages 210 to 269 shows the per cents of correct responses given to each of the 275 questions in each examination. The second column shows the order of difficulty of the items within each part, and the third column shows the order of difficulty of all the 275 questions in each test. Table 40 summarizes the facts of column 1 for all four of the new-type tests used in the Regents experiment. (Cf. p. 209 above.)

It is apparent from Table 40 that the questions in each part of each examination are well distributed with respect to difficulty. Some of the items are so easy that 90 to 100% of the students answered them correctly, and some are so difficult that correct answers were given to them by less than 10% of the students. The sampling of students whose responses were used to determine the difficulties of these items include approximately equal numbers of second- and of third-year students, and all of the fourth-year students available except in the case of French. The overlapping of classes, however, makes the distributions as to year classes of students used for this purpose relatively unimportant.

Correlation of empirical difficulties of words with Henmon and Wood frequencies. — There may be some who would question the significance of the Henmon and Wood French word list frequencies for examination purposes. The correlations of the empirical difficulties of the vocabulary items with the Henmon and Wood frequencies show conclusively that the Henmon and the Wood frequencies (appearing in columns 6 and 7 at the left of Part I of the new-type French examination) are very significant for examination purposes. The empirical difficulties of the French

vocabulary items are based on the responses of 1000 students and the correlation of these difficulty ratings with Henmon frequencies is 0.50 and with Wood frequencies 0.60. The correlation between the Henmon and the Wood frequencies for these 100 words is 0.86, and the correlation between Henmon and Wood frequencies for the 2683 words in the Wood list is 0.70.

TABLE 40

EMPIRICAL DIFFICULTIES OF INDIVIDUAL NEW-TYPE QUESTIONS. SHOWING DISTRIBUTIONS OF QUESTIONS IN EACH PART OF EACH NEW-TYPE EXAMINATION ACCORDING TO PER CENTS OF RANDOM SAMPLINGS OF STUDENTS GIVING CORRECT ANSWERS. THE FIRST COLUMN UNDER THE HEADING "FRENCH" SHOWS THAT THERE WAS ONE VOCABULARY ITEM ANSWERED CORRECTLY BY 100% OF THE STUDENTS; 29 VOCABULARY ITEMS THAT WERE ANSWERED CORRECTLY BY 90 TO 99% OF THE STUDENTS; 13 VOCABULARY ITEMS ANSWERED CORRECTLY BY 80 TO 89% OF THE STUDENTS; AND 2 VOCABULARY ITEMS ANSWERED BY 10% TO 19% OF THE STUDENTS. THE SAMPLINGS OF STUDENTS ON WHICH THE PER CENTS ARE BASED INCLUDED STUDENTS FROM ALL THREE YEAR-CLASSES IN THE LANGUAGES; THE SIZES OF THE SAMPLINGS WERE AS FOLLOWS: FRENCH, 1000; SPANISH, 400; GERMAN, 300; PHYSICS, 200.

PER CENTS OF STUDENTS GIVING CORRECT RESPONSES	FRENCH			SPANISH			GERMAN			PHYSICS
	I Vocab.	II Read.	III Gram.	I Vocab.	II Read.	III Gram.	I Vocab.	II Read.	III Gram.	
100	1			1	2			3		
90–99	29	34	10	37	36	11	49	29	8	7
80	13	20	11	25	21	8	21	18	7	23
70	8	8	22	15	6	16	14	12	15	22
60	13	8	11	11	5	17	10	6	13	24
50	14	1	12	4	3	7	5	4	11	32
40	10	2	11	4	2	6		2	15	14
30	6	1	8			13			15	14
20	4		5	2		15	1		7	6
10	2	1	6	1		5		1	8	2
0–9			4			2			1	

Reliability of orders of difficulty. — It is a generally accepted principle that in so far as possible the items or questions in an examination should be arranged in the order of difficulty. If the items are arranged in an order of difficulty, then, other things being equal, each student will be more likely than otherwise to answer all of the questions that he could answer within the time allowed. If the questions are not arranged in an order of difficulty, students may fail to answer some questions that they could answer correctly because they have wasted time trying to figure out difficult items that are beyond their powers and that came early in the test. It was for this reason, among others, that care was taken in the construction of the new-type tests to arrange the items within each part in the approximate order of their empirical difficulties. The reader will

notice that while in general the items that come first in each part are easier than the items which come later, according to the figures in column 1, there are many small, and some large, variations or disagreements between the order of difficulty and the order of appearance of the items in the tests as printed. These disagreements are due partly to mechanical errors in the set-up of the examinations for printing, but mainly to the unreliability of the order of difficulty of the items as initially established by experimental methods. The orders of difficulty in which the items were actually arranged in the Regents new-type examinations were in some cases based on the responses of less than 200 students. The disagreements between the orders of difficulty in column 2 and the orders in which the items were actually printed, show that the original determinations of empirical difficulties were not as reliable as had been hoped. It therefore becomes important to know more exactly than we know at present how many cases are required in order to establish acceptable and reliable orders of difficulty. It was for this purpose that the correlations of Table 41 were calculated.

TABLE 41

RELIABILITIES OF ORDERS OF DIFFICULTY. AVERAGES OF CORRELATIONS BETWEEN EMPIRICAL DIFFICULTIES OF THE ITEMS IN EACH PART OF THE NEW-TYPE REGENTS EXAMINATION IN FRENCH OF JUNE, 1925, BASED ON PER CENTS OF CORRECT RESPONSES IN GROUPS OF 100, 200, 300 AND 500 STUDENTS.

Regents New-Type Examination in French of June, 1925

	PART I				PART II				PART III			
N	100	200	300	500	100	200	300	500	100	200	300	500
100	.838				.896				.909			
200		.905				.958				.940		
300			.931				.954				.963	
500				.936				.964				.975
No. of correlations averaged	3	2	1	1	3	2	1	1	3	2	1	1

Table 41 shows that the reliabilities of orders of difficulty vary both with respect to numbers of students on whose responses the orders were based and with respect to the character or form of the questions. Thus the reliability of an order of difficulty based on the responses of 100 students is 0.838 for the vocabulary items of Part I of the French test, 0.896 for the true-false reading items of Part II and 0.909 for the grammar-

completion items of Part III. The reliabilities of the orders for all three parts increase gradually and fairly regularly as we increase the numbers of students on whose responses the orders were based, but the differences between the parts are not eliminated by increase in the number of cases up to 500. The reliability of an order of difficulty based on 500 cases is 0.936 for the 100 items of Part I, 0.964 for the 75 items of Part II, and 0.975 for the 100 items of Part III. The general conclusion is that orders of difficulty based on responses of 500 students are sufficiently reliable for all practical purposes, although larger numbers of cases should be used for vocabulary items wherever possible.

Validity of individual new-type questions. — The validity of any question is measured by the degree to which it correlates with acceptable criteria of achievement. If poor students answer a question correctly as often as, or oftener than, good students answer it, the question is obviously not valid. In other words if a question uniformly gives more credit points to second-year students than it gives to fourth-year students, it hinders rather than promotes the purposes of the examination. The fourth column of figures at the left of the new-type examinations reproduced above on pages 209 to 282 shows for each question the difference between the per cent of fourth-year students and the per cent of second-year students that answered the question correctly.[1]

Most of these differences are positive, showing that larger proportions of fourth-year, or highest-quarter, students answered the questions correctly than of second-year, or lowest-quarter, students. A few of the differences are negative showing that larger proportions of second-year, or lowest-quarter, students answered the questions correctly than of fourth-year, or highest-quarter, students. Earlier in this report it was stated that no items were used which did not give positive correlation with acceptable criteria of achievement. Table 42, which summarizes the facts presented in column 4, shows that 23, or a little over 2%, of the 969 questions in the four new-type examinations used in this experiment show negative differences; that is, give more credits to poor students than to good students.

The inclusion of questions showing negative differences was due in part to those accidents and mechanical mistakes which seem inevitable in handling large masses of data, and in part to the unreliability of the original determinations of the validities of the items used. The number of "bad"[2] questions is so small in relation to the total number of questions in the examinations that they had only a negligible influence on the scores of the students. Thus there were only 9 questions of 275 in the

[1] In the case of the Physics examination the difference shown is between the per cent of highest-quarter students and the per cent of the lowest-quarter students in the class who answered the question correctly.

[2] Cf. above pp. 87 for definitions of "good" and "bad" questions. The discussion here closely parallels the treatment of the validity of individual questions in the junior high school French and Spanish tests, pp. 87 to 91, above.

German examination showing zero or negative differences; and an equal proportion in the French examination. There were only 3 such questions in the 275 items in the Spanish examination and only two in the 144 items in the physics test. As nearly 98% of all the items of the four

TABLE 42

VALIDITIES OF INDIVIDUAL NEW-TYPE QUESTIONS. SHOWING DISTRIBUTIONS OF QUESTIONS IN THE FRENCH, SPANISH AND GERMAN EXAMINATIONS ACCORDING TO DIFFERENCES BETWEEN PER CENTS OF FOURTH-YEAR AND PER CENTS OF SECOND-YEAR STUDENTS ANSWERING CORRECTLY, AND IN THE PHYSICS TEST ACCORDING TO DIFFERENCES BETWEEN PER CENTS OF HIGHEST-QUARTER AND PER CENTS OF LOWEST-QUARTER STUDENTS ANSWERING CORRECTLY. THE COLUMN AT THE LEFT INDICATES DIFFERENCES BETWEEN PER CENTS OF STUDENTS IN THE FOURTH YEAR MODERN LANGUAGE CLASSES ANSWERING CORRECTLY AND PER CENTS OF STUDENTS IN THE SECOND-YEAR CLASSES ANSWERING CORRECTLY, AND BETWEEN PER CENTS OF HIGHEST-QUARTER PHYSICS STUDENTS ANSWERING CORRECTLY AND PER CENTS OF LOWEST-QUARTER PHYSICS STUDENTS ANSWERING CORRECTLY. THE NUMBERS OF CASES ON WHICH THESE DISTRIBUTIONS ARE BASED ARE THE SAME AS IN TABLE 40.

% CORRECT ANSWERS IN 4TH YEAR MINUS 2ND YEAR CLASS	FRENCH			SPANISH			GERMAN			PHYSICS
	I Vocab.	II Read.	III Gram.	I Vocab.	II Read.	III Gram.	I Vocab.	II Read.	III Gram.	
90–99				1						
80			1			1				
70	3		4				1			
60	9	1	7	4	1	5			1	
50	14	2	6	8		5	6	2	1	6
40	15	9	20	13	3	24	6	4	5	21
30	14	8	21	16	12	24	18	12	17	32
20	11	20	18	21	10	22	16	7	30	37
10	15	14	12	25	39	13	13	21	28	39
5–9	6	11	5	8	10	3	14	15	12	6
1–4	9	6	5	2		2	23	10	3	1
0	1	1							2	1
−1–4	1	1				1	2	2	1	
−5–9	1		1	1				2		1
−10	1	1								
−20				1						
−30		1								

examinations operated in the right direction and more than 82% of all the questions show positive differences of 10 or more, the percentage of efficiency is fairly acceptable.

Need for empirical verification of measurement values of individual questions. — The data on "goodness"[1] or validity of individual questions afford striking evidence of the need for great care in the construction of examinations which are designed to give accurate and meaningful informa-

[1] See above pp. 87–91.

tion about defined achievements of students. That a few "bad" questions were included even in these new-type tests, in spite of consistent appeal to objective experimental evidence on their difficulties and validities, indicates that in depending on a committee of three or four notable scholars to produce in one afternoon an examination which is to serve as a state-wide standardizing instrument, we have seriously underestimated the task.

It should be stated that our treatment of the validity of individual questions is necessarily very incomplete in this report. In order that a question be really good it must not only differentiate between second-year and fourth-year classes, or between lowest- and highest-quarter students; it must progressively give higher and higher per cents of correct responses as we go from first- to fourth-year classes, or from lowest- to highest-quarters of student groups, whether in the same or in different year-classes. In other words, the per cents of correct answers should increase from lowest to highest classes as indicated in the curves at the top of Chart 47, without irregularities or inversions at any point.

The vast majority of the questions in the new-type tests do give curves like those at the top of Chart 47; but, in addition to the 23 questions already mentioned which give curves like those at the bottom of this chart, there are a few which, in spite of showing positive differences as between second- and fourth-year classes, show negative differences as between second- and third-, or between third- and fourth-year classes. At first thought it seems impossible that any reasonable question could be framed which would be correctly answered by mediocre oftener than by superior students; yet many questions of both the old and the new types behave in this curious manner. Very little is known as to the precise cause or causes of such anomalies, but they undoubtedly lie in subtle ambiguities in the form, content or context of the questions, — ambiguities which mislead superior students but which are too fine to be noticed by students who are still in the elementary stages or are dull or both. It is to enable those interested to study the characteristics of both good and bad questions, with a view to learning how to cultivate the former and avoid the latter, that we have indicated the "order of goodness" of all the questions in each of the new-type tests used in this experiment. This order of goodness is indicated in the fifth column of figures at the left of the questions in the tests reproduced on pages 210 to 282.

The reader may find considerable amusement, and perhaps profit in indulging in a guessing game with the figures in the third and fifth columns at the left of the questions in the new-type tests. He may, for example, learn something of his ability to judge the difficulty of questions by selecting any two items in one of the tests, deciding which is the more

difficult, and comparing the decision with the figures in column 3. Or he may guess at the relative differentiating powers, i.e. validities, of two questions and compare his judgments with the figures in column 5.

The easiest of the 275 items in the new-type French test is *faire*, No. 2 in Part I, and the hardest is No. 100 in Part III; the best question is No. 47 in Part III, and the worst question is No. 72 in Part II. A study

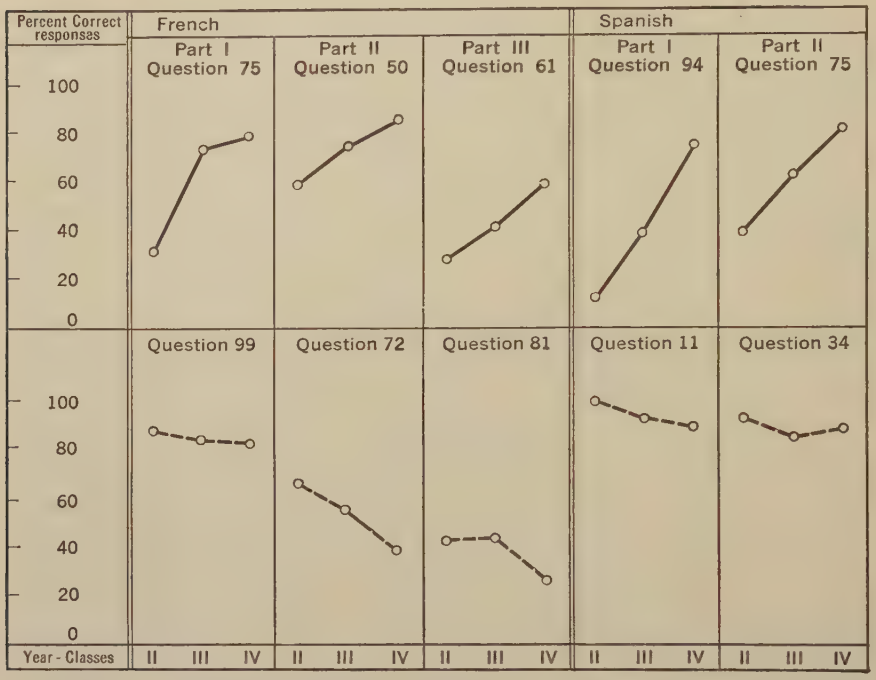
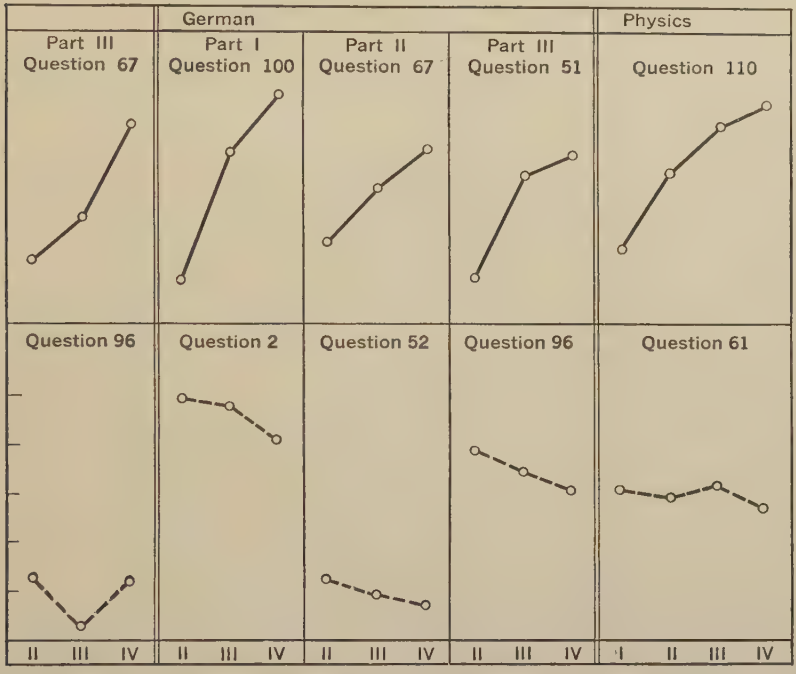


Chart 47. — Illustrating graphically the validity of individual questions in the new-type examinations in French, Spanish, German and Physics. The graphs at the top of the chart relate to the good questions and those at the bottom to bad questions. The chart is to be read as follows: Item 75, in Part I of the new-type French examination, was answered correctly by 31% of French II students, by 72% of French III students, and by 79% of French IV students. The corresponding per cents for Item 99 in Part I of the French test are 86, 84 and 82%. The graphs for the questions in the other modern

of some of the bad questions will show obvious defects in them. For example, No. 79 in Part I, *sol* (= soil), contains two very tricky alternatives, *sun* and *alone*. Number 81 in Part III shows a negative difference which is probably due to a slight inadequacy in the scoring key. There are several other questions showing defects that are more or less obvious at the present writing but which escaped detection before the tests were printed. For obvious reasons, the tests are reproduced on pages 210 to 282 exactly as they were presented to the students in June, 1925, as parts

of the Regents examinations. The identification of the easiest and hardest, best and poorest questions in the Spanish, the German and the Physics tests is left to the interest of the reader. (Cf. p. 209 above.)

Validity of individual old-type questions. — The three most heavily weighted questions in the old-type modern language examinations are (1) the foreign language-to-English translation, (2) the English-to-foreign



language tests are to be read in the same way. The graphs at the extreme right show that question 110 of the physics examination was answered correctly by 29% of the lowest-quarter students, by 62% of the second lowest-quarter students, by 80% of the third-quarter students, and by 89% of the highest-quarter students; and that question 61 was answered correctly by 62% of the lowest-quarter students and by 55% of the highest-quarter students.

language translation, and (3) the free composition questions. The intercorrelations of these three questions in the French II and Spanish II examinations, and also their correlations with total new-type scores, are shown in Table 43.

The English-to-foreign language translation question gives the highest correlation with new-type scores in both the French II and the Spanish II examinations; the lowest correlation with new-type scores is given by the free composition question in French II and by the Spanish-to-English

translation question in the Spanish II examination. The intercorrelations of the three French II questions average about 0.54 and of the three Spanish II questions about 0.37. When we recall that these three questions in the old-type papers account for about two-thirds of each old-type examination, their intercorrelations seem much too low. It seems probable to us that while the lowness of the relationships is partly due

TABLE 43

CORRELATIONS OF TOTAL NEW-TYPE EXAMINATION SCORES WITH THE OLD-TYPE TRANSLATION AND FREE COMPOSITION QUESTIONS, AND INTERCORRELATIONS, MEANS AND SIGMAS OF THE LATTER, BASED ON RANDOM SAMPLINGS OF FRENCH II AND SPANISH II STUDENTS WHO TOOK THE REGENTS EXAMINATIONS IN JUNE, 1925. THE CORRELATIONS IN COLUMN (2) SHOW THAT THE ENGLISH-TO-FOREIGN LANGUAGE TRANSLATION QUESTIONS ARE THE MOST VALID OF THE THREE OLD-TYPE QUESTIONS IN BOTH LANGUAGES; BUT THE LARGEST NUMBERS OF CREDIT POINTS ARE ASSIGNED TO THE TRANSLATIONS IN THE OPPOSITE ORDER. THE FREE COMPOSITION QUESTIONS ARE THE MOST HEAVILY WEIGHTED, AS SHOWN BY THEIR GREATER SIGMAS, IN COLUMN (6).

(1)	(2)	(3)	(4)	(5)	(6)	(7)
OLD-TYPE FRENCH II EXAMINATION QUESTIONS	NEW-TYPE FRENCH EXAM.	FRENCH TO ENGLISH	ENGLISH TO FRENCH	OLD-TYPE QUESTIONS: Mean	OLD-TYPE QUESTIONS: Sigma	NUMBER OF CASES
French-to-English translation (25 credits)	0.586			19.94	3.32	1110
English-to-French translation (20 credits)	0.600	0.508		12.03	3.61	1110
Free composition (or oral credit) (20 credits)	0.460	0.542	0.588	14.01	4.15	1110
OLD-TYPE SPANISH II EXAMINATION QUESTIONS	NEW-TYPE SPANISH EXAM.	SPANISH TO ENGLISH	ENGLISH TO SPANISH	OLD-TYPE QUESTIONS Mean	OLD-TYPE QUESTIONS Sigma	NUMBER OF CASES
Spanish-to-English translation (25 credits)	0.464			20.55	2.4	971
English-to-Spanish translation (20 credits)	0.614	0.368		13.56	2.8	971
Free composition (or oral credit) (20 credits)	0.523	0.330	0.423	15.07	3.0	971

to the independence of the three functions tested by the three questions, it is largely due to the unreliability of each of the three old-type question forms. This unreliability is due primarily to the inadequate sampling of modern language materials and of student performances in such old-type questions, and partly to the subjectivity in the scoring of the students' responses. We have shown in preceding tables and charts that the vocabulary samplings of the old-type examinations are not only meagre

but are very heterogeneous with respect to difficulty or frequency of occurrence. The presence of only three or four fifth-thousand words in a passage which second-year students are asked to translate might easily cause an excellent French II student to make a very poor showing — especially when the scoring is subjective.

General weaknesses of translation and free composition questions. — The translation and free composition questions in the old-type examinations suffered also because of the general nature and indefiniteness of the abilities and skills which they measure. The translation questions in both directions require the ability to read both languages as well as to write in both languages. Therefore, in translating a foreign passage into English, the failure of the student may be due more to poor expression in English than to lack of understanding of the foreign passage, and failure to translate an English passage into the foreign language may, in multilingual communities, be due as much to difficulty in understanding the English passage as to inability to express ideas in the foreign language. The free composition question tests not only ability to express ideas in terms of the foreign language but several other more or less specific abilities which we may designate by the term usually employed, "imagination." Some modern language scholars hold that the complexity of abilities implied in old-type examinations is the prime merit of old-type forms of questions. That these abilities are all exceedingly important is undeniable, but it is a serious question whether it is the function of modern language examinations to measure them. The task of measuring simple and unadorned knowledge of the foreign language itself is a large enough task for any single examination; it is hardly wise or reasonable to expect modern language examinations, which are known to be imperfect in measuring modern language ability, to measure also general intelligence, imagination and spiritual well-being. These qualities are just as significant for other departments of instruction as for modern languages; moreover, they are important enough to merit separate and very elaborate examinations. The attempt to measure these general qualities in terms of specific subject matter examinations is the main cause of the general meaninglessness of most school marks. Some French teachers may feel entirely justified in encouraging a gifted and imaginative youth by giving him a high grade in French in spite of poor achievement in mastering the language itself; or in flunking a youth who knows the language well but who lacks imaginative fire and aesthetic perspective; but the main effect of such confusion of meaning of grades is to make the intelligent educational guidance of students almost impossible. The fundamental purpose of school grades is to convey in universally meaningful terms, accurate information concerning *defined* achievements and traits. Examinations given in elementary modern language courses should confine themselves

to the task of describing accurately and meaningfully the student's mastery of the language itself; to attempt to go further than this in such examinations is an ambitious assumption of a duty which belongs to other parts of the educational system of the state.

Fallacy of substituting oral credit for free composition questions. — A more specific example of the indefinite meaning of grades based on old-type examinations is found in the dual character of the free composition question. According to the Regents rules, oral credit, as determined by the local schools, may be substituted for the full value of the free composition question. Out of a little over 1100 old-type French II papers used in deriving the figures of Table 43 more than 600 showed that oral credit of from 16 to 19 points had been substituted for the possible 20 points of credit on the free composition question. No paper was found in the 1100 in which the full 20 points of possible credit on the free composition question was given for oral work.

It is well known that the oral and aural skills may be relatively independent of achievement in the written language. It is often asserted by those entitled to an opinion that the oral-aural work has materially improved in New York State and elsewhere since the introduction of oral and aural examinations. A student might be able to write a fairly respectable composition in French of 150 or 200 words without being able to pronounce correctly all or a majority of the words that he has written. In determining oral-aural credit some teachers base their marks almost wholly on mechanical perfection of pronunciation, while others put considerable emphasis on content in the student's foreign language speech. Thus the grade on the old-type Regents examination is a variable measure of a variable complex of dependent and of independent abilities and skills. Whenever oral credit is substituted for the free composition question on a paper, it means that the direct application of the Department readers' standards is impossible for a fifth of the total possible credit on a paper. The variable emphasis which teachers give to oral work mentioned above is responsible to no small extent for the variations in individual school standards, which, as described earlier in this report, remain partially uncorrected by the Department reviewing.

Special weaknesses of free composition question. — There is no form of question and no aspect of a foreign language for which the direct application of centralized reviewing is more needed than for the free composition question and the oral-aural skills. In reading free compositions and in judging oral-aural skills the subjective judgment of the rater has a freedom which is practically beyond restriction in the absence of directed centralized reviewing. It is unfortunate that the beneficent influence of the State Department readers in reducing the variable standards of the local schools is least effective in that part of the modern language examina-

tion in which such reviewing is most needed. The freedom of the teacher to substitute oral credit for the free composition question is probably a large factor in producing the low correlations for the free composition question displayed in Table 43.

The special weaknesses of the free composition question are too well known to need comment. If the topics assigned are kept within the range of what could be reasonably expected of students, such as, "A Day in the Country," "A Trip to the Museum," "A Voyage," "My Family," "My Home," "The School Room," etc., the students may be unfairly coached. On the other hand, if in choosing topics the examiners go beyond the range of coaching they may easily go beyond the range of the students' foreign language vocabulary or beyond his experience, or beyond both.

It has been argued by some that the new-type tests are of little value because they do not measure free composition ability and the oral-aural skills. The answer to this is that they do measure these abilities and skills indirectly as accurately as, and probably more accurately than, the traditional tests measure them. But, admitting the charge that the new-type forms of tests do not measure these things, the only conclusion that can be drawn is that we must search for means of measuring speech and hearing factors as accurately as the new-type tests measure ability with the written language. Since the two kinds of abilities are relatively independent, our measures of them should be kept separate in any case. To declare that the new type is of little value because it does not measure oral-aural skills is like throwing away a dollar because it is not two dollars; also, to fail a student in a French course merely because he cannot pronounce would be just as confusing as to report an over-weight child as under-weight merely because he is under the average height for his age.[1]

Irrelevant activities in old- and new-type examinations. — The fundamental reason for the inferiority of the old-type examinations is that they involve so many irrelevant activities that only a fraction of the examination is really expended in displaying the type of achievement which the examination is supposed to measure. The fact that everything in the student's answers must be written out in legible longhand in itself places strict and narrow limits on the samplings of modern language materials which examiners may include in their examinations. Moreover, the activity of writing in longhand involves many opportunities for purely mechanical mistakes, and is a potent distraction to the student which disturbs him even in those moments when he is actually not moving his pen. Translation questions are particularly inefficient mirrors of modern language achievement because in addition to the weaknesses just mentioned they are not flexible enough to permit of detailed adjustments of size and

[1] Cf. pp. 96 to 98, above, on the limitations of the new-type tests.

TABLE 44

NUMBERS OF TIMES THE SAME WORD IS USED IN OLD- AND IN NEW-TYPE EXAMINATIONS. NO ACCOUNT OF THE FREE COMPOSITION QUESTIONS IS TAKEN IN THIS TABLE; NOR ARE THE ENGLISH-TO-FRENCH TRANSLATION QUESTIONS ACCOUNTED FOR EXCEPT IN THE CASE OF THE REGENTS 1925 EXAMINATIONS.

	No. of Times Same Word Is Used	Numbers of Words in Old-Type Regents Examinations											
		1924						1925					
		French II		French III		French IV		French II		French III		French IV	
		N	%	N	%	N	%	N	%	N	%	N	%
	1	182	67.0	140	67.0			116	67.0	120	72.0	111	73.0
	2–4	56	21.0	53	26.0			41	24.0	32	19.0	31	20.0
	5 +	32	12.0	15	7.0			16	9.0	16	9.0	11	7.0
Total Different Words (a)		271		208				170		166		153	
Total Discourse (b)		703		450				376		364		280	
Ratio a/b			39.0		46.0				45.0		46.0		55.0
		College Entrance Examination Board Examinations											
	1	136	66.0	161	69.0	171	71.0	157	66.0	177	70.0	190	71.0
	2–4	48	24.0	48	20.0	45	19.0	54	23.0	54	21.0	54	20.0
	5 +	21	10.0	23	11.0	25	10.0	25	11.0	22	9.0	24	9.0
Total Different Words (a)		206		235		241		237		253		270	
Total Discourse (b)		462		585		557		563		603		649	
Ratio a/b			45.0		40.0		43.0		42.0		42.0		42.0

		Junior High School New-Type[1] 1925				Regents New-Type[1] 1925			
		Part II Reading	Part III Grammar	I + II + III		Part II Reading	Part III Grammar	I + II + III	
	1	251	59	313	59.0	238	44	329	72.0
	2–4	159	14	174	33.0	75	20	95	21.0
	5 +	30	4	42	8.0	29	4	35	7.0
Different Words (a)		442	75	529		344	68	459	
Discourse (b)		1289	125	1517		915	141	1157	
Ratio a/b		34.0	60.0	35.0		38.0	48.0		48.0
		57.0	79.0			69.0	65.0		
		36.0	19.0			22.0	29.0		
		7.0	2.0			9.0	6.0		

[1] Part I of the new-type test consists of 100 French words each of which appears only once.

character of vocabulary and other samplings and because the activity of translating itself involves special kinds of skills and abilities not particularly relevant to genuine modern language teaching goals. Translating may be a good form of discipline, but in view of the fact that so few of our modern language students become translators, surely more directly useful ways of disciplining young minds can be found, especially since so large a proportion of modern language students never go beyond the second-year class. Translation questions necessarily involve repetitions of words and phrases, many of which have little if any measurement value.

Table 44 shows that an average of about 30% of the different root words in both the old- and the new-type examinations are used twice or oftener in the same or in inflected forms. In the old-type examinations nearly 10% of the different root words are used five times or oftener, while only about 7.5% of the different root words in the new-type tests are repeated five or more times. Even if the number of repetitions were greater in the new-type than in the old-type the waste of time and energy would still be greater in the old-type examinations because in them every word, whether repeated or not, has to be written out, whereas in the new-type they have only to be read. Thus in the old-type three-hour examinations the student has to write at least 600 words of running discourse in order to expose himself to a sampling of about 225 different root words; while in the new-type 90-minute examinations the student has to write only 175 digits or symbols and about 125 words in order to react to about 500 different words. In some of the old-type examinations words like *le* and *de* have to be written out 40 or 50 times each, although they apparently have real measurement value in only five or six of the contexts in which they occur. The wastes of repeating the same words and the distraction of writing everything down in longhand are inherent in the old-type form of examination. Even with the utmost care, examiners will probably never be able to reduce such wastes greatly, as long as they confine themselves to old-type question forms. It is unnecessary to repeat here that the old-type examinations referred to in Table 44 were constructed with extraordinary care by examiners of many years' experience in constructing old-type examinations.

VI

COSTS OF OLD- AND OF NEW-TYPE EXAMINATIONS

On account of the sample method of reviewing Regents examinations by the Department readers, and the impossibility of finding the exact number of papers which are actually reviewed, we are unable to compare the costs of new- and old-type Regents examinations used in this experiment. The annual reports of the Secretary of the College Entrance Examination Board include exhaustive statistics on the total and distributed costs of examinations in each subject matter. Since the Regents and College Board examinations are very similar in method and carefulness of construction and scoring, we believe that a comparison of College Board costs with new-type Regents costs will be a fair substitute for a comparison between the costs of the new- and old-type Regents examinations. With this in view, Table 45, showing relative costs of old- and new-type examinations in French, Spanish, German and Physics, has been prepared. The lowest per-paper reading cost for the old-type modern language examinations shown in Table 45 is about eight times larger than the highest per-paper cost of the new-type. Experience has shown that the cost of constructing new-type examinations would be about the same as that of constructing old-type papers under the present system, and that the cost of printing new-type papers will not average above one cent per paper over a series of years.

The actual cost of the State Department reviewing of old-type papers is, of course, considerably lower than is indicated by the figures of Table 45. This is due to the fact that the State Department readers do not review every paper sent in by the schools; they read varying proportions of the papers from different schools and accept the school ratings on the remainder of the papers. The data presented in previous sections of this report indicate that the sample method of reviewing is a fundamental weakness in the Regents system. It therefore becomes of interest to learn what the cost would be if the State Department readers should give every Regents paper (e.g. in French) a genuine independent reading uninfluenced by knowledge of the school rating. There were nearly 21,000 papers in French II, III and IV in June, 1925. According to column 3 in Table 45 the number of French papers read per hour is 2.4. Calling this average 3 for the sake of avoiding a fraction and dividing it into 21,000, it appears that 7000 working hours or 1120 working days would be neces-

sary to score the Regents French examinations of June, 1925, if the latter were as long as the College Entrance Board French papers. Putting it in terms of dollars, the cost of scoring 20,000 French papers at the lowest rate reported by the College Entrance Examination Board would be over

TABLE 45

RELATIVE COSTS OF SCORING OLD- AND NEW-TYPE EXAMINATIONS IN FRENCH, SPANISH, GERMAN AND PHYSICS. COLUMN (1) SHOWS THE AVERAGE OF THE LOWEST "PER-PAPER COSTS OF READING" REPORTED BY THE COLLEGE ENTRANCE EXAMINATION BOARD FOR THE YEARS 1920, 1922 AND 1925 IN THE RESPECTIVE SUBJECT MATTERS. COLUMN (2) SHOWS THE AVERAGES OF THE HIGHEST "PER-PAPER COSTS" REPORTED BY THE BOARD FOR THE SAME YEARS. THESE FIGURES ARE FOR "READING" ONLY, EXCLUSIVE OF ALL OTHER EXPENSES INCLUDING SUPERVISION. THE FIGURES FOR THE NEW-TYPE SCORING COSTS IN COLUMN (4) INCLUDE TECHNICAL SUPERVISION, CHECKING ALL SCORING AND COUNTING, AND TRANSFERRING SCORES TO THE REGENTS BOOKLETS. OVER THREE-SEVENTHS OF THE COST OF SCORING THE NEW-TYPE MODERN LANGUAGE EXAMINATIONS IS ACCOUNTED FOR BY PART III, THE GRAMMAR-COMPLETION TEST. THE COST OF PRINTING NEW-TYPE PAPERS WILL NOT AVERAGE ABOVE ONE CENT PER PAPER OVER A SERIES OF YEARS. THE COST OF CONSTRUCTING NEW-TYPE EXAMINATIONS OVER A SERIES OF YEARS WILL BE ABOUT THE SAME AS THE COST OF CONSTRUCTING OLD-TYPE UNDER PRESENT CONDITIONS. COLUMN (3) SHOWS THE AVERAGE NUMBER OF PAPERS READ PER HOUR BY THE AVERAGE COLLEGE BOARD READER.

SUBJECTS	COLLEGE ENTRANCE EXAMINATION BOARD			NEW-TYPE
	Per-paper reading-costs; Averages of years 1920, 1922, 1925		Number of papers read per hour	Per-paper cost of scoring
	Low	High		
Columns	(1)	(2)	(3)	(4)
French	$0.71	$0.92	2.4	$0.07
Spanish	0.54	1.07	3.0	.07
German	1.11	1.73	1.7	.07
Averages for Modern Languages	$0.78	$1.24	2.37	$0.07
Physics	$0.43	$0.43	4.1	$0.03

$14,000; while a similar number of new-type French papers could be scored for $1400 or less. These differences are too large to need comment. The important finding of this report, however, is that regardless of relative costs, the objective test furnishes the only known means by which the State Department can do that for which the Regents examinations were instituted, namely, furnish the state educational system with accurate measures of defined achievements expressed in comparable units and reckoned from stable and uniform standards.

VII

CONCLUSIONS AND RECOMMENDATIONS

Space limitations do not allow us to describe the Regents system in detail. In earlier sections certain features of the Regents system have been mentioned, notably in connection with Table 21 on pages 133 to 136 above. The background and significance of the present experiment will be more adequately appreciated if the general features of the Regents system are understood, especially since so many erroneous notions concerning the Regents examinations are entertained even by teachers within New York State. For this purpose we shall quote excerpts from an article appearing in "The University of the State of New York," *Bulletin to the Schools*, Vol. 12, No. 2, Oct. 1, 1925.

History and methods of the Regents examinations. — "Regents academic examinations were first given in 1878 in response to concerted action by the academic principals of the State who urged the Board of Regents to establish a series of outside objective tests in secondary subjects for the purpose of measuring the work of their schools and of establishing proper standards of scholarship. Until 1906 these examinations were made in the Regents office, under the direction of the secretary of the Board, by a staff of inspectors and examiners, small at first but gradually enlarged as the schools grew and the use of the examinations became more widespread. . . .

"In 1906 . . . the Regents of the University created a State Examinations Board to serve as a liaison agency between the schools and the central administration of the University. . . .

"This Board has twenty members: fifteen representing the colleges, superintendents and secondary schools, and five representing the Department. Ever since the day of its establishment each revision of our courses of study has been made upon its recommendation and by committees of its selection. Similarly, the examinations based upon these syllabuses have been prepared under its direction by committees of teachers chosen from various types of high schools and colleges in the State.

". . . The making of Regents examinations is not, then, merely a matter of office routine. They represent the collective thinking of the teachers of the State. In each major field of study, such as history or English, a committee, usually of four members, of which only one is a member of the State Department, prepares the question papers in that field. These

papers are therefore an expression of the experience and judgment of teachers actually engaged in teaching or in supervision. On those question committees the Department has had only minority representation and has sought to have associated in the performance of these highly important duties the strongest men and women in the field of collegiate and of secondary education.

"The papers thus prepared by the several committees are subjected to critical examination for form by the editor of examinations and are then submitted for final revision to a committee, usually of seven members, appointed by the State Examination Board. Regents question papers are therefore the product of thoughtful consideration, over a considerable period of time, by competent teachers experienced in the classroom or in supervision.

"It is doubtful whether a more democratic scheme of coöperation in the preparation of Regents examinations could be devised. . . .

"There are two respects in which this scheme of statewide tests is more flexible and has created higher standards; namely, (1) the preparation of papers by the committee system in which there is a full participation of those actually engaged in instruction, and (2) the development both in the schools and in the Department of a standardized scheme for rating.

". . . The board and its committees have not been unmindful of the new theories respecting educational measurements. It is constantly seeking methods of improving the examination system. It is the desire of the board and of the Department to adopt new measures whenever results warrant, as is evidenced by recent changes in the form of certain papers and by experiments with a new type of test for which greater accuracy and greater ease of rating are claimed. As illustrative of this may be cited certain changes in the form of papers set in English, the incorporation of a silent reading test, the inclusion of true-false, recognition and completion questions as parts of papers in other subjects and recasting of algebra papers to provide a simpler mechanical arrangement for more rapid rating.

"With the improvements in the form and in the content of examinations which have been brought about in recent years, there has come an equally important improvement in the method of rating answer papers. In most of the larger schools departmental organization is such that answer papers in a given subject group may be rated by the committee system. This in large measure eliminates the personal equation that sometimes makes the work of the individual reader unsatisfactory. The committee system also obtains in the work of the examining staff of the State Department particularly in the summer when the force of readers is largely augmented. The quality of the personnel engaged in reading papers at the State Education Building has been greatly strengthened. During the summer over

160 experienced teachers are called into the service to assist the permanent examiners. These are college graduates with evidence of sound judgment and with a minimum teaching experience of three years in the subject which they examine. During these years, by a careful process of selection and of testing there has been built up a staff of summer examiners of unusual merit representing in each field a group of the best teachers in the State. . . .

". . . In New York City the examinations are used for supervisory purposes in order to reveal points of strength or weakness in instruction both in the schools as a whole and in separate groups. . . .

"Regents examinations have played an important part in the educational history of the State and while they have sometimes been adversely criticised, such criticism usually has been made by persons uninformed or biased. Generally these tests have been regarded as an important factor in maintaining proper educational standards not only in the large city high schools but in the small high schools in rural communities."

The Regents system sound in principle. — To some readers who have carefully digested the data of this report, it may appear that the preceding excerpts fall somewhat short of giving a complete description or a fair appraisal of the actual workings of the Regents system. However the reader may judge the merits of the appraisal implied in these excerpts, the writer is convinced that, all things considered, the Regents system is fundamentally sound *in principle.* In preceding pages of this report we have exposed by means of objective evidence the inadequacies of the examinations used by the Regents and have presented a brief analysis of what seemed to be the main causes of these inadequacies. But it cannot be too strongly emphasized that these weaknesses arise from the particular forms of examinations used and from certain details of procedure, none of which are *necessarily inherent* in the Regents system. In other words, we find no evidence from any source which calls in question the underlying principles of the Regents system. The purpose of examinations is to give accurate measures or meaningful descriptions of defined achievements and to express such measures in terms which can be understood by all competent educators, to the end that such information may be used constructively in guiding students educationally and in administering the educational system to the best advantage of all concerned. In order to fulfil this purpose it is necessary that all the educational measurements within a given educational system be expressed in terms of comparable units reckoned from uniform and stable standards or points of reference.

Destructive criticisms of Regents system not justified by facts. — The Regents system of examinations, like that of the College Entrance Examination Board, has been severely criticized in some quarters as being wrong

in principle, reactionary, autocratic, technically faulty and extravagantly costly. Dr. Kruse's report on "The State's System of Examinations" in *The Rural School Survey of New York State*, Volume 1, Part 5, is a noteworthy example of extremely unfavorable criticism.

The present report lends support to only one of Dr. Kruse's conclusions, namely, that the examinations used are technically faulty. We find no evidence supporting his other conclusions, but on the contrary find that his recommendations looking towards abandonment or weakening of the Regents system are unsound and destructive rather than constructive. The old-type examinations are admittedly defective and unworthy both of the confidence reposed in them and of the system in which they are used. But the principle of uniform centrally administered examinations is sound and is indispensable to the wise administration of our educational system. Every charge which has been made against the Regents examinations holds with multiplied force against locally derived examinations which Dr. Kruse and others seem to recommend as a substitute. There can be no doubt about the existence of grave imperfections in the Regents and College Entrance Examination Board examinations; but to abandon them in favor of locally derived examinations would merely exaggerate existing imperfections and add many others without retaining any of the good features which the carefully derived and administered Regents and College Board examinations possess.

Duties of examining agencies include research and information services. — Nor is there any danger essential in centrally administered systems of examinations so long as progress is not forestalled by the creation of vested interests and autocratic authority. It seems that sensitiveness to new conditions and to the demands of sound progress are assured in the Regents system by the distribution of authority and by the democratic composition of the various advisory boards and committees that are responsible for the construction and supervision of the Regents examinations in the various subject matters. To some advocates it may seem that the centralized examining agencies, including the Regents, have failed to keep pace with the progress which research scholars have made in recent years in improving instruments for the measurement of mental capacities and educational achievements. The first duty of executive officers in charge of examining agencies is to administer the available instruments which command the confidence of the educational system. But a fully coördinate duty is that of improving available instruments and of evaluating improvements suggested by research workers everywhere, and of keeping their educational constituencies adequately informed about such improvements. One aspect of this latter duty is a continuous and aggressive exposure, and a fearless publication, of whatever defects exist in the examinations and procedures that are in force

and that command the confidence of the public at any given time. An official in charge of an examining agency who has failed either to ascertain and publish the weaknesses of the examinations administered, or who has failed to inform himself accurately and precisely about the scientific evidence back of suggested improvements in examining agencies has been guilty of a serious dereliction to duty. That such charges do not apply to the Regents system is evidenced by the large number of experiments, including the present one, that have been undertaken by the Regents examiners in recent years in efforts to verify and to profit by the values of new types of examinations. The experiment with which this report is concerned is only one of many bits of evidence to the effect that those in charge of the Regents examinations are willing to expose the weaknesses of the examinations used and are anxious to adopt any proposed modifications, the values of which for the Regents system can be scientifically and fairly demonstrated.

The time-serving conception of achievement is indefensible. — If the objective and standardized tests are adopted for the modern languages it will be possible to define year-classes or credits in terms of actual achievement rather than in terms of length of time spent "taking" modern language work. It is due to an exaggeration of the importance of clock time spent in class as a condition for — and measure of — achievement, and to a lack of appreciation of individual differences, that we have so long reposed confidence in a system of separate subjective examinations for each year-class; and that the incredible overlapping of classes and misplacement of students, exposed in this report, have continued so long undetected and uncorrected. The retention in the French II class of the 300-odd students who in actual achievement are above the fourth-year average is a clear example of the evil workings of the old time-serving conception of education which has persistently ignored the fact of individual differences. It is only one of many examples that might be cited wherein the rules of examining agencies put a premium on stupidity and laziness and a penalty on intelligence and industry. Regents rules, like those of the College Entrance Examination Board, prejudge every modern language student without trial by jury or any other of the amenities which the law accords to the worst criminals, and sentences them to a fixed number of years in modern language classrooms; until the student has done time to the full extent of his sentence he is not even allowed to *try* the advanced examinations except by rare and very special dispensations. Some readers may find it difficult to believe that such a situation could actually exist in the most advanced educational system of the twentieth century, but such worthy doubts are sadly dispelled when we read in the caption of the Regents French II examination[1] that, "The

[1] See above pp. 199–200.

minimum time requirement (for admission to the French Two-Year examination) is five recitations a week for a school year in (*a*) first-year French, (*b*) second-year French," and similar statements from the captions of the third- and fourth-year examinations. This rule is no better, and in all its insidious effects is almost certainly worse, than a rule designed to regulate by chronological ages the heights of school desks and the sizes of shoes and of hats that school children might wear. In view of the near equality of the second-, third- and fourth-year old-type examinations as to difficulty, one might surmise that these rules are set up precisely because the examining authorities fear that some second-year students might surreptitiously pass a fourth-year examination if allowed to try it. If this motive has had any part in leading to the adoption of the rule, it is a remarkable acknowledgment of the weakness of the types of examinations used by such agencies. It would be much more reasonable to give extra promotion to students who learned more in less time than the average. If a student is to study a modern language at all, he should be allowed to study it in a class of his peers, regardless of whether he has "had" five 40-minute periods per week for one or for four years. This can be done by defining both year-classes and individual student achievements in terms of scores of such standardized, highly reliable, and valid tests as we have in the new-type parts of the Regents examinations used in this experiment. Thus the French classes might be defined as follows: at the beginning of the school year in September, 1926, the French II class should include only students who in June, 1926, secured scores between 90 and 130 on the Regents new-type French examination; the French III class should include only students who secured scores between 130 and 170; and the French IV class should begin with students whose scores range from 170 to 200. If these classes are divided into semester groups, the score limits should be as follows: French II A 90 to 110, French II B 110 to 130; French III A 130 to 150, French III B 150 to 170; French IV A 170 to 185; French IV B 185 to 200. Students scoring above 200 in June, 1926, should receive credit in French IV, regardless of how long they have studied French; if they continue the study of French after thus receiving credit for French IV, they should take work more advanced than that in regular French IV classes. The score-limits here suggested are merely illustrative. The important thing is that if these definitions of classes should be adopted, every competent teacher everywhere and at any time would understand them, so long as these or similar comparable tests were used.

Summary. — The general conclusions from the data of this experiment are that the new-type examinations are roughly twice as reliable and valid as the old-type examinations of equal time allowance; that the new-type examinations afford comparable measures for all classes in a given subject

matter in the same and in different years and thus offer a means of eliminating overlapping of classes and variations in local school standards to a much greater extent than they are eliminated by the old-type Regents examinations; and that the new-type tests over a series of years will cost not more than 10% as much as old-type examinations, as administered and read by the College Entrance Examination Board, cost.

The validity of the comparison of new- and old-type costs on the basis of College Entrance Board figures cannot be denied on the ground that the actual cost of Regents reviewing under the present system is very much less than the cost reported by the College Entrance Examination Board, because the data reported above show conclusively that the sample method of reviewing cannot for a moment be considered as adequate. There is considerable evidence showing that the results of the College Entrance Board Examinations evince significant variations in standards and inaccuracies in spite of the careful and costly way in which the examinations are read. That such variations and inaccuracies are necessarily greater when the reading is less complete is self-evident.

In this study we have discussed the weaknesses existing in the Regents examinations and procedures for purely constructive ends. Our fundamental recommendation, therefore, is that the Regents system should take advantage of the values of the new-type without surrendering any of the *real* values of the old-type examinations. The committees charged with the task of constructing old-type questions should be provided with means for securing objective and experimental evidence on the difficulties and measurement values of old-type questions. It seems almost certain that the full values of the old-type question forms have never been fully exploited because of the reliance in their construction on the subjective opinion of scholars rather than upon experimental evidence.

It seems fair to suggest that the new-type forms of questions be adopted for half of each Regents examination period. More detailed recommendations obviously are not within the province of this report. How and by whom the new-type tests, if adopted, will be constructed; how, when and by whom they shall be scored; what weight shall be given to the new-type parts of the examinations; in how many and in what subject-matters shall the new-type forms be tried out first; what budgetary redistributions and what administrative reorganizations shall be necessary in the Examinations and Inspections Division of the State Education Department, — these are only a few of the important questions that would have to be faced if the new-type forms were adopted by the Regents. They are obviously questions which can be solved only by careful consideration on the part of the officers of the State Department of Education.

PART THREE

SECOND SURVEY OF MODERN LANGUAGE ACHIEVEMENT IN THE JUNIOR HIGH SCHOOLS OF NEW YORK CITY

SECOND SURVEY[1] OF MODERN LANGUAGE ACHIEVEMENT IN THE JUNIOR HIGH SCHOOLS OF NEW YORK CITY

Introduction. — In June, 1925, Form A of The American Council Beta French test was administered to approximately 19,000 students of French, and Form A of The American Council Beta Spanish test to approximately 6500 students of Spanish in the junior high schools of New York City. In June, 1926, Forms B of The American Council Beta French and Spanish tests were administered to 18,870 students of French and to 3940 students of Spanish in the junior high schools. The results of the 1925 tests have been analyzed and reported above.[2] This paper relates to the 1926 test results, and is offered as a preliminary report on a few of the most important aspects of the data at hand, so that the readers of the report on the 1925 data may have the "follow-up" data in the same volume.

TABLE 46

NUMBERS OF NEW YORK CITY JUNIOR HIGH SCHOOL STUDENTS TAKING THE AMERICAN COUNCIL BETA TESTS IN FRENCH AND SPANISH IN 1925 AND 1926.

CLASS	FRENCH		SPANISH	
	June, 1925 Form A	June, 1926 Form B	June, 1925 Form A	June, 1926 Form B
8A	3,736	4,054	1,486	1,239
8B	3,041	4,072	1,332	945
9A	2,604	2,177	658	489
9B	2,417	1,723	688	426
RB	2,766	2,842	850	333
RC	2,415	2,360	683	275
RD	1,898	1,642	721	233
Totals	18,877	18,870	6,418	3,940

Number of junior high school students taking the tests in 1925 and in 1926. — Table 46 shows the numbers of students taking each test in each year by semester classes. It would have been interesting to study the variations in numbers of students in the same classes in the two years,

[1] This study was made possible by grants from The Commonwealth Fund, and from The Carnegie Corporation through The Modern Foreign Language Study; by the coöperation of the research staff in the office of Dean Hawkes of Columbia College; and by the coöperation of Mr. Jacob Greenberg, Director of Foreign Languages in the Junior High Schools of New York City. Especial thanks are due, and heartily made, to Professors R. H. Fife, Algernon Coleman, V. A. C. Henmon, J. P. W. Crawford, C. M. Purin, and their associates in The Modern Foreign Language Study for support and helpful counsel.

[2] Pp. 3 to 103, above.

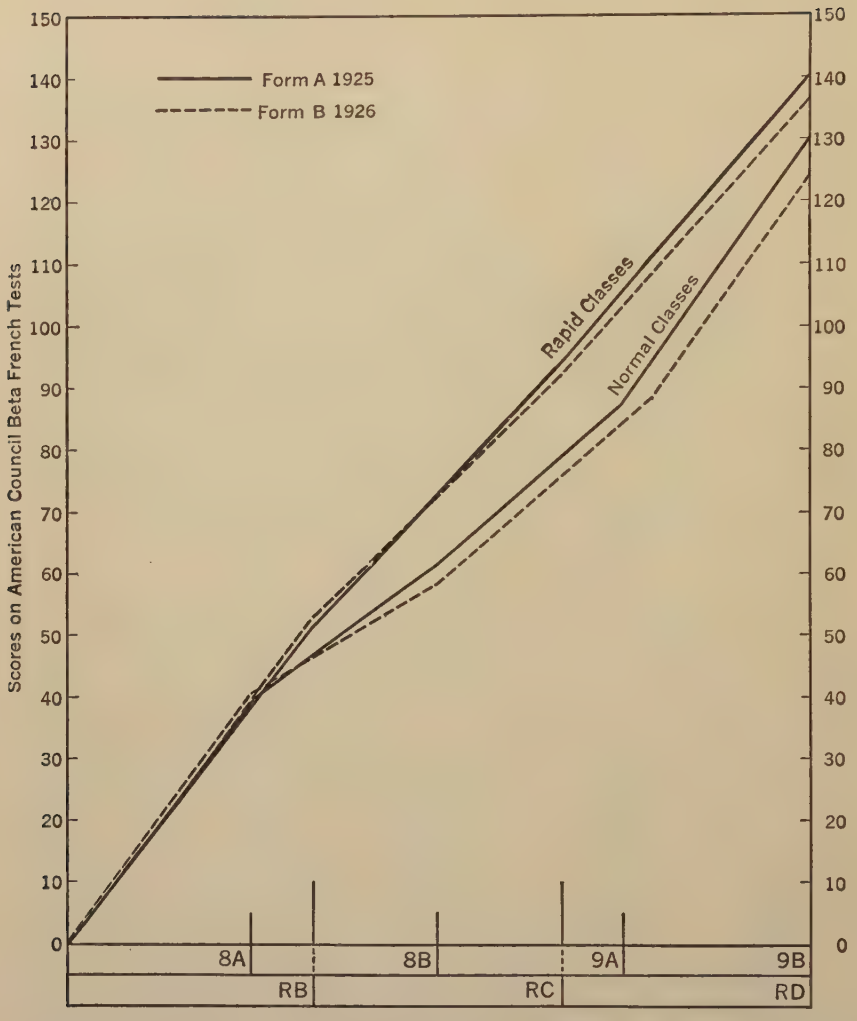


CHART 48. — Graphs of 1925 and 1926 French class medians, from Table 47, showing equivalence of Forms A and B of the American Council Beta French Test. The near-rectilinearity of the lines indicates that both Forms of the test are adequately scaled in difficulty throughout the whole range of achievement, and that the units of the scale are approximately equal throughout the whole range of achievement in French in the junior high schools of New York City. The slight downward dip in the curves for the normal classes is easily accounted for by the retention of "repeaters" and by the promotion out of 8A of some students who should be eliminated or held in 8A.

but the means for going into this important question were lacking. Why, for example, should there be only 3000 students in French 8B in 1925 and 4000 in 1926? Still more important, why should there be a falling off of enrolment in Spanish from 6400 to 3940? Such variations are too

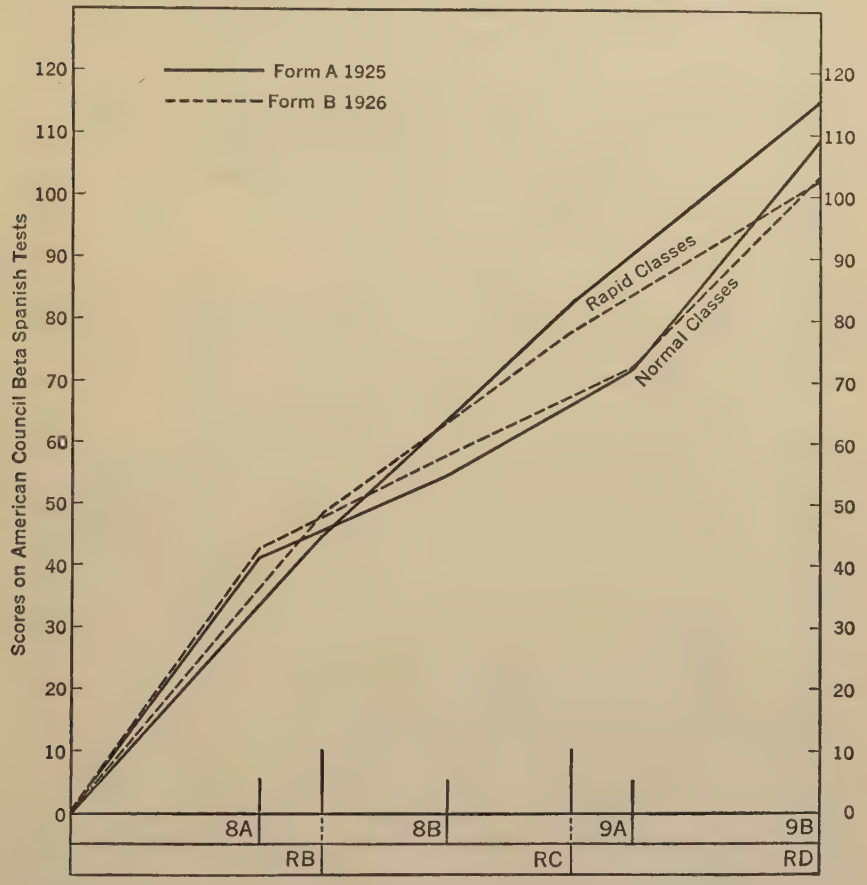


CHART 49. — Graphs of 1925 and 1926 Spanish class medians, showing equivalence of the American Council Beta Spanish Tests. See Chart 48. The apparent differences are very likely due to changes in the character of the student body in Spanish which occurred between June, 1925 and June, 1926: the Spanish enrolment in June, 1926, was less than 62 per cent. as large as that in June, 1925.

large to be due to chance and are therefore worthy of careful analysis. Whatever caused the landslide of nearly 40 per cent of the 1925 Spanish enrolment out of Spanish in 1926 is peculiar to Spanish since it did not affect the French enrolment.

Equivalence of Forms A and B. — The American Council Beta tests in French and Spanish were originally constructed in two equivalent Forms,

A and B. Unless the modern language tests that are used in our schools are either equivalent or comparable, the measurements derived from different tests will be more or less meaningless and therefore of little use in the administration of modern language work and in research looking toward its improvement. The equivalence of Forms A and B of the American Council Beta tests was determined experimentally as explained

TABLE 47

EQUIVALENCE OF FORMS A AND B OF AMERICAN COUNCIL BETA TESTS; SHOWING MEDIANS AND QUARTILES OF EACH CLASS IN FRENCH AND SPANISH. THE NUMBER OF CASES ON WHICH EACH MEDIAN IS BASED IS SHOWN IN TABLE 46.

		FRENCH		SPANISH	
		Form A 1925	Form B 1926	Form A 1925	Form B 1926
8A	Lower Quartile	29.2	32.7	32.8	33.3
	Median	39.5	40.4	41.3	42.8
	Upper Quartile	50.0	49.2	49.7	54.2
8B	Lower Quartile	47.0	47.6	44.7	47.6
	Median	61.0	58.3	54.7	57.7
	Upper Quartile	77.7	70.5	66.7	70.0
9A	Lower Quartile	70.5	67.0	60.5	58.3
	Median	87.3	83.6	71.8	72.0
	Upper Quartile	105.5	100.0	86.0	86.8
9B	Lower Quartile	110.2	99.6	86.4	82.5
	Median	130.0	123.8	109.1	103.1
	Upper Quartile	148.0	140.7	136.3	126.0
RB	Lower Quartile	40.0	42.4	35.0	40.0
	Median	51.2	52.8	44.9	48.6
	Upper Quartile	64.3	64.1	53.2	57.4
RC	Lower Quartile	75.5	75.1	68.6	65.6
	Median	94.0	91.8	82.6	78.0
	Upper Quartile	111.5	110.0	98.0	92.0
RD	Lower Quartile	116.7	115.0	93.8	88.0
	Median	140.2	136.2	115.3	102.5
	Upper Quartile	160.5	154.0	134.2	122.5

above on pages 3 to 4. Table 47 and Charts 48 and 49 justify the claims of equivalence made in that report, since the differences for the French test are entirely negligible and the slightly larger differences for the Spanish test may easily be explained in terms of the small numbers of cases involved — especially in the norms for Form B. (We have already called attention to the fact that the 1926 Spanish enrolment is less than 62% of the 1925 enrolment).

In both charts the near-rectilinearity of the graphs of the norms is worthy of note. The Form A line for the rapid advancement French

classes is practically straight throughout; and the Form B line parallels it with only negligible deviations. Aside from confirming the almost exact equivalence of the tests these lines indicate that the tests are adequately scaled in difficulty for the whole range of achievement in the junior high schools and that the units of the scale are approximately equal throughout the whole range of achievement. In both charts the downward dip of the curves for the normal classes is easily accounted for by the retention of "repeaters" and by the promotion out of 8A of some students who should be eliminated from modern language work or held in 8A. The most important indication of Charts 48 and 49 is that the experimental method of constructing tests, if used with sufficient care, will produce tests capable of affording comparable measurements of modern language achievement.

Influence of homogeneity of classes on achievement. — Two conditions disclosed by the 1925 data were the excessive heterogeneity of classes within a given school and the excessive variability of standards in different schools. It was hoped that at least some of the schools might use the results of the 1925 tests to reclassify the students of French and Spanish and thus secure homogeneous classes. If this had been done we could have compared the progress of homogeneous classes with that of heterogeneous classes; but, as we shall see, so few of the schools found it possible to use the 1925 test results to any considerable extent for reclassifying purposes that no careful study of the influence of homogeneity on achievement could be made. However, correlations between degree of homogeneity and amount of achievement were calculated on the theory that, although the variations in homogeneity were apparently due entirely to chance, nevertheless some relation might be found between homogeneity and achievement. These correlations turned out to be almost exactly zero. This is by no means evidence that homogeneity is undesirable, for even if its influence were very great, it would have been — and in this case most probably was — masked by still more powerful influences inherent in each individual school situation. We shall illustrate concretely the effects of some of these hidden influences in succeeding sections.

Constancy of progress. — Table 48 shows the correlations between the scores achieved by students on Form A in 1925 and on Form B in 1926 of the French and Spanish tests. Under continuously ideal school conditions, it is plausible to assume that normally students will progress at a uniform rate, so that if a student in his first year achieves what is normal for the first-year students we should expect him in his second year to achieve what is normal for second-year students; also, if in his first year he got within the highest or lowest 10% of first-year students we would expect him, in his second year, to be in the highest or lowest 10% of second-year students. Since continuously ideal school conditions, as far

as we know, do not exist, it is impossible to verify this plausible theory. Common sense favors the theory, but however reasonable or unreasonable it may seem, Table 48 shows that a uniform rate of progress in individual students in modern language work is the exception rather than the rule in the junior high schools of New York City.

TABLE 48

CONSTANCY OF PROGRESS IN MODERN LANGUAGE WORK. CORRELATIONS BETWEEN SCORES ON FORM A TAKEN IN JUNE, 1925, AND SCORES ON FORM B, TAKEN IN JUNE, 1926. COLUMNS 1 AND 6 SHOW THE CORRELATION COEFFICIENTS; COLUMNS 2 AND 7 THE COEFFICIENTS OF ALIENATION, BY THE FORMULA $k=\sqrt{1-r^2}$; 3 AND 8 THE SIGMAS OF THE FORM B 1926 SCORES; 4 AND 9 THE STANDARD ERRORS OF ESTIMATE OF FORM B 1926 SCORES FROM FORM A 1925 SCORES; AND COLUMNS 5 AND 10 SHOW THE NUMBER OF CASES INVOLVED IN EACH CORRELATION IN COLUMNS 1 AND 6.

	FRENCH					SPANISH				
	1	2	3	4	5	6	7	8	9	10
	r_{25-26}	K	σ_{26}	$\sigma_{26.25}$	N	r_{25-26}	K	σ_{26}	$\sigma_{26.25}$	N
$8A_{25}$–$9A_{26}$	.326	.942	24	22.6	1533	.553	.832	25.0	21.0	372
$8B_{25}$–$9B_{26}$	.455	.890	29	26.0	1382	.628	.777	34.0	26.0	342
RB_{25}–RD_{26}	.463	.885	28	25.0	1338	.389	.920	24.5	22.5	208

The correlations are all very low; the coefficients of alienation in columns 2 and 7 of Table 48 show that the relation between the achievements of students in successive years, *as the classes were organized and taught in the sessions of 1924–1926*, is very nearly one of pure chance. There are only slight differences between the three classes described at the extreme left of Table 48, but the order of magnitude of the correlations from highest to lowest, according to columns 4 and 9, is the same for both French and Spanish, — 8A–9A, RB–RD, 8B–9B. Columns 4 and 9 show the magnitude of the correlations for each language in terms of identical units, that is, in terms of standard error of estimate expressed in terms of the score-points of each test. If there is any tendency toward a uniform rate of progress in individual students these correlations demonstrate that in the junior high schools of New York City that tendency is almost completely overridden by more powerful forces.

That such powerful influences are due to chance variations in the pedagogical conditions in individual schools is strongly indicated by the correlations of Table 49. Table 48 shows that in the junior high schools of New York City a uniform rate of achievement for individual students is the exception, and Table 49 shows that a uniform rate of achievement for classes composed of the same individual students is equally exceptional.

TABLE 49

CONSTANCY OF PROGRESS IN MODERN LANGUAGE WORK. CORRELATIONS BETWEEN 1925 MEDIANS AND 1926 MEDIANS OF INDIVIDUAL SCHOOL CLASSES MADE UP OF THE SAME INDIVIDUAL STUDENTS.

	CORRELATIONS BETWEEN MEDIANS OF		
	8A, 1925 and 9A, 1926	8B, 1925 and 9B, 1926	RB, 1925 and RD, 1926
French	0.361	0.222	0.494
No. of pairs of Medians	(31)	(27)	(27)
Spanish	0.322	0.701	0.000
No. of pairs of Medians	(14)	(12)	(8)

The method of calculating the correlations of Table 49 will be illustrated by a consideration of the first coefficient in the table, 0.361: Among the junior high schools of New York City there were 31 that had 8A classes taking the examination in June, 1925, and 9A classes taking the examination in June, 1926. These 8A and 9A class rolls were compared and a list was made for each of the 31 schools of the individual students common to both classes. These lists included both the Form A 1925 and Form B 1926 scores, and the medians of the two sets of scores for each list of students were calculated. Thus we secured 31 pairs of medians, the correlation between which turned out to be 0.361.

The magnitude of the correlations in Table 49 is so small as to constitute a serious indictment of the system of educational guidance, or lack of it, which is apparently common to all departments of instruction in our educational system. The average of the correlations of Table 49 is only 0.35, which means that the relation between the rate or amount of progress of a whole class in its first year of modern language work and the rate or amount of progress of the same group of individuals in the same school in the second year of its modern language work is very nearly a pure chance relation, *under the present organization of modern language work in the junior high schools of New York City.* The low correlations of Table 48 might plausibly be explained as due to more or less accidental shifts in the interests of individual students, or to temporary or permanent variation in their ability to learn a modern language, or to any one of a dozen other things which might affect the achievement of an individual student; but it is highly improbable that any one or all of these chance influences could operate in the same direction on each and every individual in a class of 25 to 150 students. The variable rate of achievement of whole classes can be explained only in terms of variations in the learning conditions in individual classrooms, such as the assignment of a class that has for the first year

enjoyed a good teacher to a poor teacher, or the reverse; or the shift from a good textbook or a good method of instruction to a bad textbook, or bad method of instruction, or the reverse; or the merging of a hitherto very homogeneous class with a heterogeneous group of students, or the reverse. The operation of any one or of all of these possible influences, with the general background of extreme heterogeneity of classes and of incredible variations in the standards of individual schools, is sufficient to explain the failure of some schools to develop demonstrated ability in large groups of students, and the failure of other schools to make manifest the real ability of other large groups of students until the second year.

TABLE 50

NUMBERS AND PROPORTIONS OF STUDENTS WHO SCORED LOWER ON FORM B IN JUNE, 1926, THAN ON FORM A IN JUNE, 1925.

		TOTAL NUMBER OF STUDENTS WHO TOOK BOTH FORMS OF TEST AND WHOSE PAPERS COULD BE MATCHED, THAT IS, WHO TOOK FORM A IN 1925 IN, *e.g.* 8A, AND FORM B IN 9A.	SCORED LOWER IN 1926 THAN IN 1925, HAVING BEEN REGULARLY IN CLASS DURING THE INTERVENING YEAR AND PROMOTED AT LEAST ONCE, *e.g.* 8A TO 8B.	
			Numbers	Per cents
French	8A–9A	1533	33	2.2%
	8B–9B	1382	12	0.8%
	RB–RD	1338	3	0.2%
Spanish	8A–9A	372	12	3.4%
	8B–9B	340	6	2.2%
	RB–RD	208	2	1.0%
Total French and Spanish		5173	68	1.3%

It would be almost absurd to try to explain the small magnitude of the correlations of either Table 48 or Table 49 as due to any considerable extent to the unreliability of the American Council Beta tests. The report on the analysis of the 1925 data is sufficient to eliminate this argument.[1] Incidentally the correlation scatter-diagrams of Table 48 enable us to secure additional evidence on the reliability of the tests. According to Table 50, only 68 out of 5173 students secured lower scores on Form B in 1926 than they secured on Form A in 1925, and of these 68 less than a dozen were lower by more than five or ten points in a total possible range of scores of 9 to 220. In comparison with similar data from equivalent tests that have been given a year apart, these figures are quite gratifying. The American Council Beta tests are not, of course, *perfectly* reliable, and some of the variations pointed out are undoubtedly due to this fact; but

[1] Cf. above pp. 3 to 103.

it is easy to conceive other reasons why at least 13 out of 1000 students tested a year apart might secure lower scores on the second than on the first test.

Extent of reclassification.—As noted above it was hoped that the objective and comparable measurements secured by means of the American Council Beta tests in 1925 would be used by some of the schools for purposes of making their classes homogeneous and of equating the standards for the various semester groups in all of the schools. Table 51 furnishes an indication of the extent to which homogeneity of classes was achieved.

Table 51 shows that the interquartile ranges of the classes are exceedingly variable. One of the classes has an interquartile range of only 8 or 9

TABLE 51

COMPARISON OF VARIABILITIES OF CLASSES IN INDIVIDUAL SCHOOLS. DISTRIBUTIONS OF INTERQUARTILE RANGES OF FRENCH AND SPANISH CLASSES IN THE JUNIOR HIGH SCHOOLS OF NEW YORK CITY IN JUNE, 1925, AND JUNE, 1926. THESE FIGURES ARE BASED ONLY ON SCORES OF STUDENTS WHO WERE COMMON TO BOTH CLASSES INDICATED ABOVE EACH PAIR OF COLUMNS. FORM A WAS USED IN 1925 AND FORM B IN 1926 OF BOTH FRENCH AND SPANISH TESTS.

INTERQUARTILE RANGES IN TERMS OF SCORE-POINTS	FRENCH						SPANISH					
	8A–9A		8B–9B		RB–RD		8A–9A		8B–9B		RB–RD	
	A 1925	B 1926	A 1925	B 1926	A 1925	B 1926	A 1925	B 1926	A 1925	B 1926	A 1925	B 1926
8– 9											1	1
10–11			1		1		2					
12–13	5		1	1	2		3		2		1	
14–15	4		2	2	1	1	3		3	1	1	
16–17	6	3	5	2		2	3	1			1	
18	3	1	5	1	8	2		6	2	2	3	
20	6	3	3	1	2	2	1		1	2		
22	1	4	2	2	5	4	1	1	1	1	1	1
24	2	7	1	1	2	2		1		1		2
26		2		3		5	1					
28	3	2	2	3	2	2		3				
30		4	2	4	1	3			1	1		2
32		1	1	2	2	2		1				2
34		2		1					1	1		
36						1		1		1		
38			1							1		
40				1								
42		1										
44				2								
46												
48												
50										1		
52–53									1			
No. of schools	30	30	26	26	26	26	14	14	12	12	8	8
Median interquartile range	18	25	20	28	21	26	15	21	19	24	18	28

points, and one has an interquartile range of over 50 points. Fifty points represents a greater range of achievement than the difference between the 8A and 9A city-wide class medians, and is greater than the difference between the medians of any pair of successive semester classes in either language. Ten of the 82 individual school French classes in June, 1925, had interquartile ranges of less than 14 score-points, and 14 had interquartile ranges of 28 or more points. In June, 1926, only one French class had an interquartile range of less than 14 points and 31 individual school classes had interquartile ranges greater than 28 points. These figures are based only on scores of students who were in the same school and in corresponding classes in 1925 and 1926. This illustrates strikingly the need for continuous educational guidance of students.

Table 51 shows that some of the junior high schools started the year with comparatively well-organized and homogeneous classes and kept their classes homogeneous throughout the year; while most of the schools started badly and became worse so far as homogeneity of classes is concerned. The methods used by the former should be carefully studied and made known to the latter schools. It cannot be too often urged that more attention must be given to the task of *learning* students and of *guiding* them; we are now emphasizing *teaching* at the expense of *good* teaching because a large fraction of the teaching effort is misguided and wasted. In this connection the reader may review the argument of pages 99 to 103 of this volume.

Variable standards of schools.—Table 52, showing distributions of the medians of the same individual school classes whose interquartile ranges are distributed in Table 51, further emphasizes the need for continuous constructive guidance of students on the basis of objective and comparable measurements of defined achievement.

Table 52 gives us a picture of the persistence of the incredibly large differences in standards of individual schools disclosed in the report of the 1925 test results.[1] In spite of the availability of the accurate and comparable Form A measurements of June, 1925, two schools started the session of 1925–1926 with classes which were called "8B" but which were actually above the 8B average. One of these two classes in September, 1925, was just average for 9A classes (see Table 52, column headed "8A–A–1925"). The third column in Table 52 shows that in three schools the classes which started as 9A classes in September, 1925, were actually below the average of 8A classes and that two were above the average of 9A classes.

It is difficult to explain the persistence of these variations in the standards of individual schools, especially since most of the modern language teachers in the city were anxious to improve the situation. There can be little doubt, however, that the lack of flexibility in the existing administrative

[1] See above pp. 3 to 38.

TABLE 52

COMPARISON OF MEDIANS OF CLASSES IN INDIVIDUAL SCHOOLS. DISTRIBUTIONS OF MEDIANS OF FRENCH AND SPANISH CLASSES IN THE JUNIOR HIGH SCHOOLS OF NEW YORK CITY IN JUNE, 1925, AND JUNE, 1926. THESE FIGURES ARE BASED ONLY ON SCORES OF STUDENTS WHO WERE COMMON TO BOTH CLASSES INDICATED ABOVE EACH PAIR OF COLUMNS. FORM A WAS USED IN 1925 AND FORM B IN 1926 OF BOTH FRENCH AND SPANISH TESTS.

SCORE VALUES OF MEDIANS	FRENCH						SPANISH					
	8A–9A		8B–9B		RB–RD		8A–9A		8B–9B		RB–RD	
	A 1925	B 1926	A 1925	B 1926	A 1925	B 1926	A 1925	B 1926	A 1925	B 1926	A 1925	B 1926
20–24	2				1		1					
25–29	4		1		1							
30–34	6		1		2		1					
35–39	3		1				4				1	
40–44	7		2				4	1	1		3	
45–49	3	1	2		6		2		3		2	
50–54	3	1	3		5		2		1		1	
55–59	1	1	5	1	4				1		1	
60–64	1		3		3				4	1		1
65–69		2	4		3			5				
70–74		3	1		1	1		3	1			
75–79		3	2					1				
80–84	1	5		1				2				
85–89		3	1	1		1		2		1		
90–94		7				1				1		1
95–99		2		1						1		1
100–104		1	1	1		1			1	2		1
105–109				2	1					3		1
110–114		1		1		1						
115–119		1				1				1		2
120–124				5		1				1		1
125–129				5		1						
130–134				3		4						
135–139				3		3						
140–144						2						
145–149						5						
150–154				1		2						
155–159				1		2						
160–164				1		1						
165–169												
170–174												
175–179										1		
No. of schools	31	31	27	27	27	27	14	14	12	12	8	8
Upper Quartile	47.0	88.1	67.8	133.8	62.1	148.2	46.3	81.3	63.7	110.0	50.0	117.5
Median	40.5	84.5	58.5	125.5	53.5	137.5	41.3	71.7	60.0	105.0	45.0	105.0
Lower Quartile	31.5	74.6	48.4	108.4	47.3	123.7	38.7	67.5	49.1	95.0	41.6	95.0

organization is partly or largely responsible. The needs of modern education have outgrown their administrative and organizational habiliments. Within the last generation our public school system has become increasingly a matter of mass education without any corresponding increase or development of the administrative system, so far as the intelligent and effective guidance of the multiplied thousands of school children is concerned. School administration has been made comparatively efficient on the material side, but practically nothing has been achieved in the direction of keeping individual students continuously in proper relation to the curriculum. We know infinitely more about the buildings in which our children attend classes than we know about the children themselves; partly because we have spent more money and time on learning about buildings than we have spent in learning about children. Recent developments in the field of educational measurement enable us to secure relatively accurate measures of defined achievement in comparable and meaningful terms, but our administrative system is not at present able to use this information in the constructive guidance of students. *Tables 51 and 52 constitute concrete evidence of this fact.* Efficiency of teaching has been too much associated with what might be called "minor tactics" of the classroom. School administrations, in addition to caring for the physical plant and announcing the major goals of education in general terms, must cease to lean so heavily upon the *tactics* of the individual classroom and must develop a *strategy* which will give the tacticians of the classroom a real opportunity. It is certainly poor strategy to put into a 9A class of 50 students ten who are below the 8A average and ten who are above the 9B average; yet, according to both the 1925 and 1926 data, this is very nearly the typical situation in the junior high schools of New York City.

Progress of semester classes in individual schools in relation to initial achievement and to initial homogeneity.—In order to illustrate concretely some of the facts reported above in technical terms, Charts 50 to 52 have been prepared. Since these charts are exactly parallel in form, only the first one will be described in detail.

The first thing to fix in mind in regard to Chart 50 is that the heavy vertical lines at 40 and at 87 represent the approximate city-wide 8A and 9A medians. The heavy and light horizontal lines represent the interquartile ranges of the scores of the same students in 1925 and 1926, respectively. Thus the first pair of horizontal lines at the top of the chart relate to a school whose identity we have concealed under the number 1. In this school there were 58 8A students who took Form A in June, 1925, and who took Form B in 1926 as 9A students. The last pair of lines on the chart relate to 72 students from school 31 who took Form A in June, 1925, as 8A students and who took Form B in June, 1926, as 9A students. In every case students who were in the 8A class in 1925 and not in the 9A class

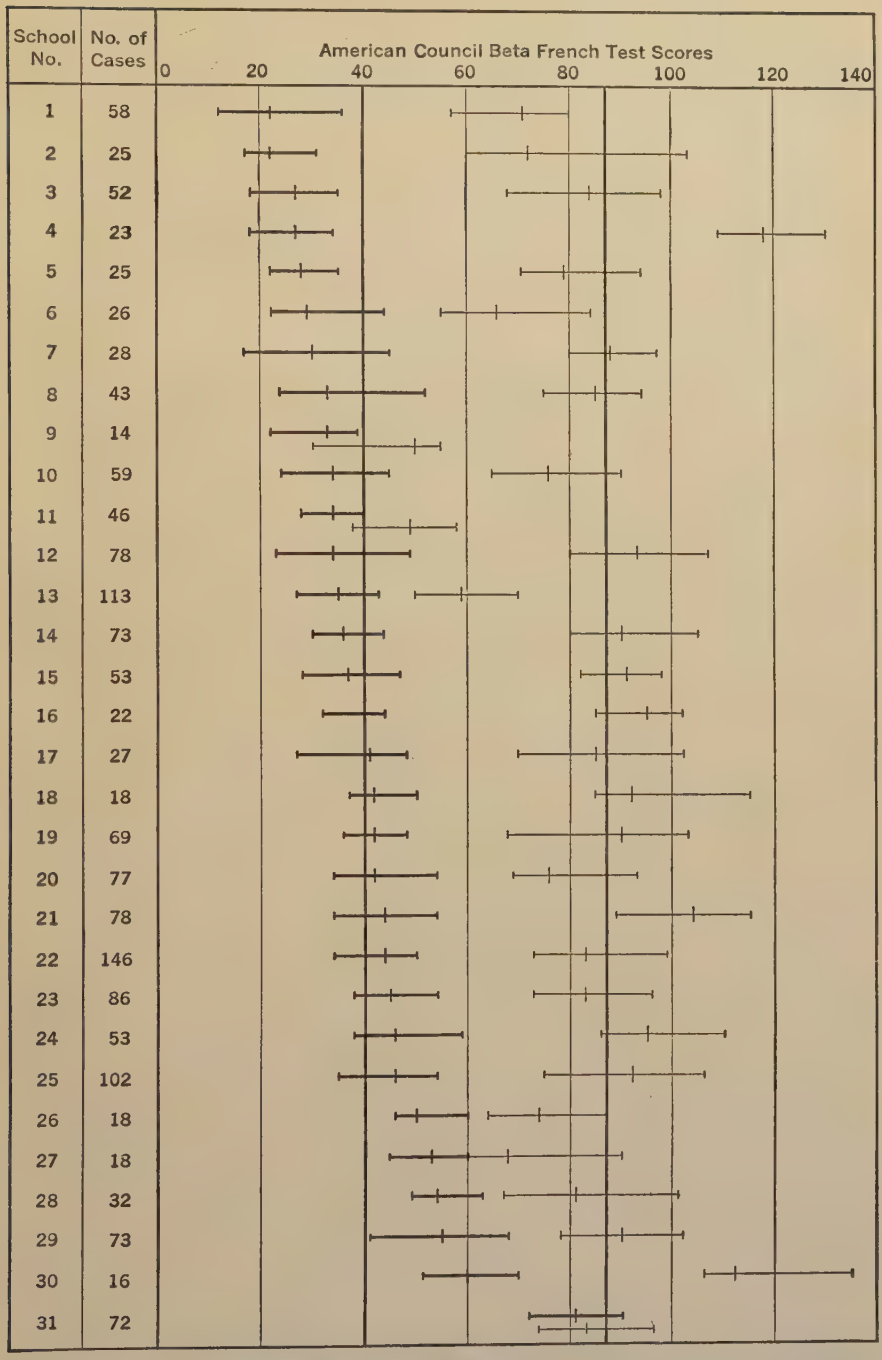


CHART 50. — Medians and interquartile ranges of American Council Beta French Test scores of identical groups of students who took Form A in June, 1925 as 8A students and Form B in June, 1926 as 9A students.

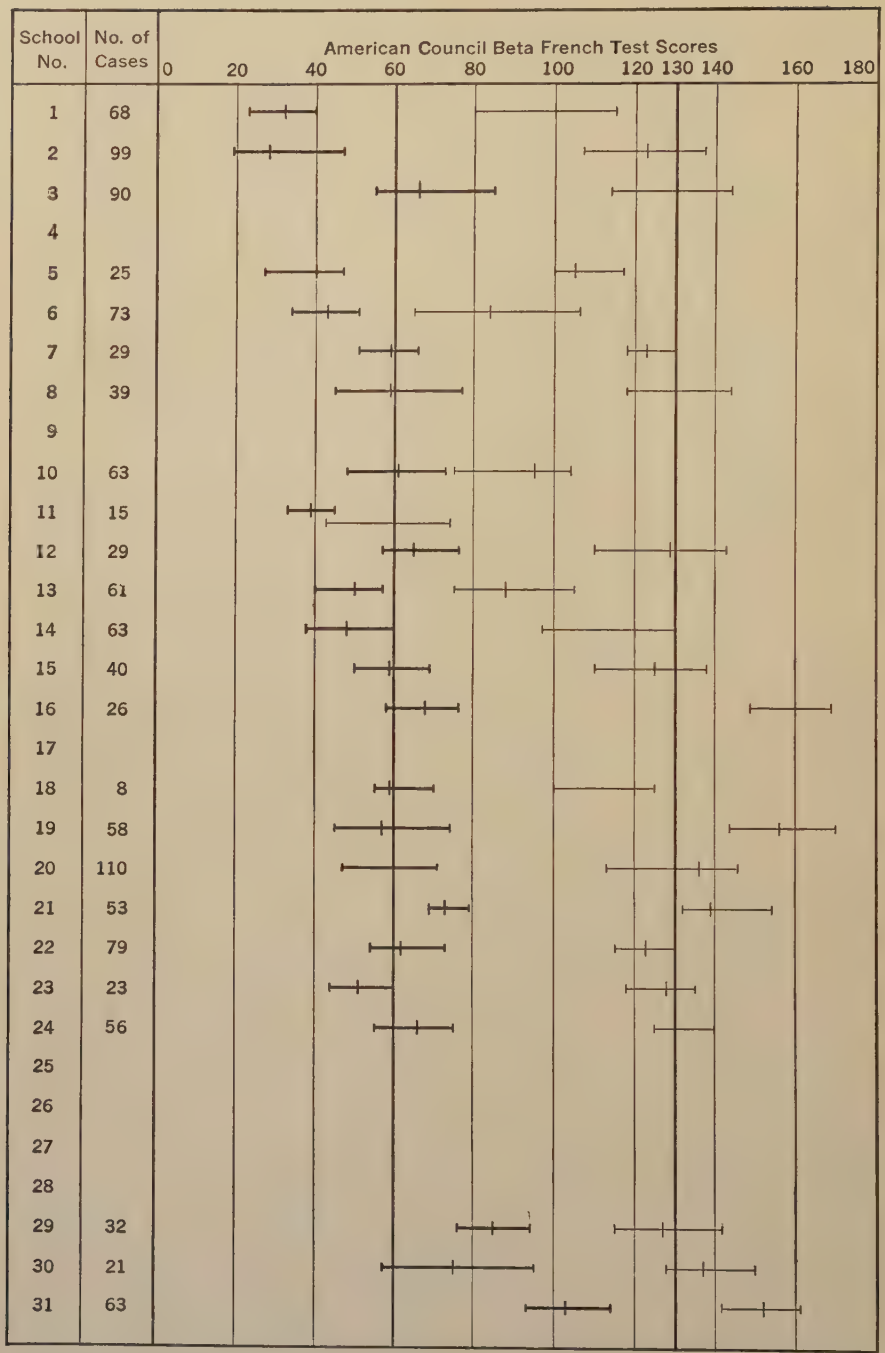


CHART 51. — Medians and interquartile ranges of American Council Beta French Test scores of identical groups of students who took Form A in June, 1925 as 8B students and Form B in June, 1926 as 9B students. The heavy vertical lines at 60 and 130 indicate the approximate city-wide 8B and 9B medians, respectively. This chart is similar to and should be read in the same way as Chart 50, *q. v.*

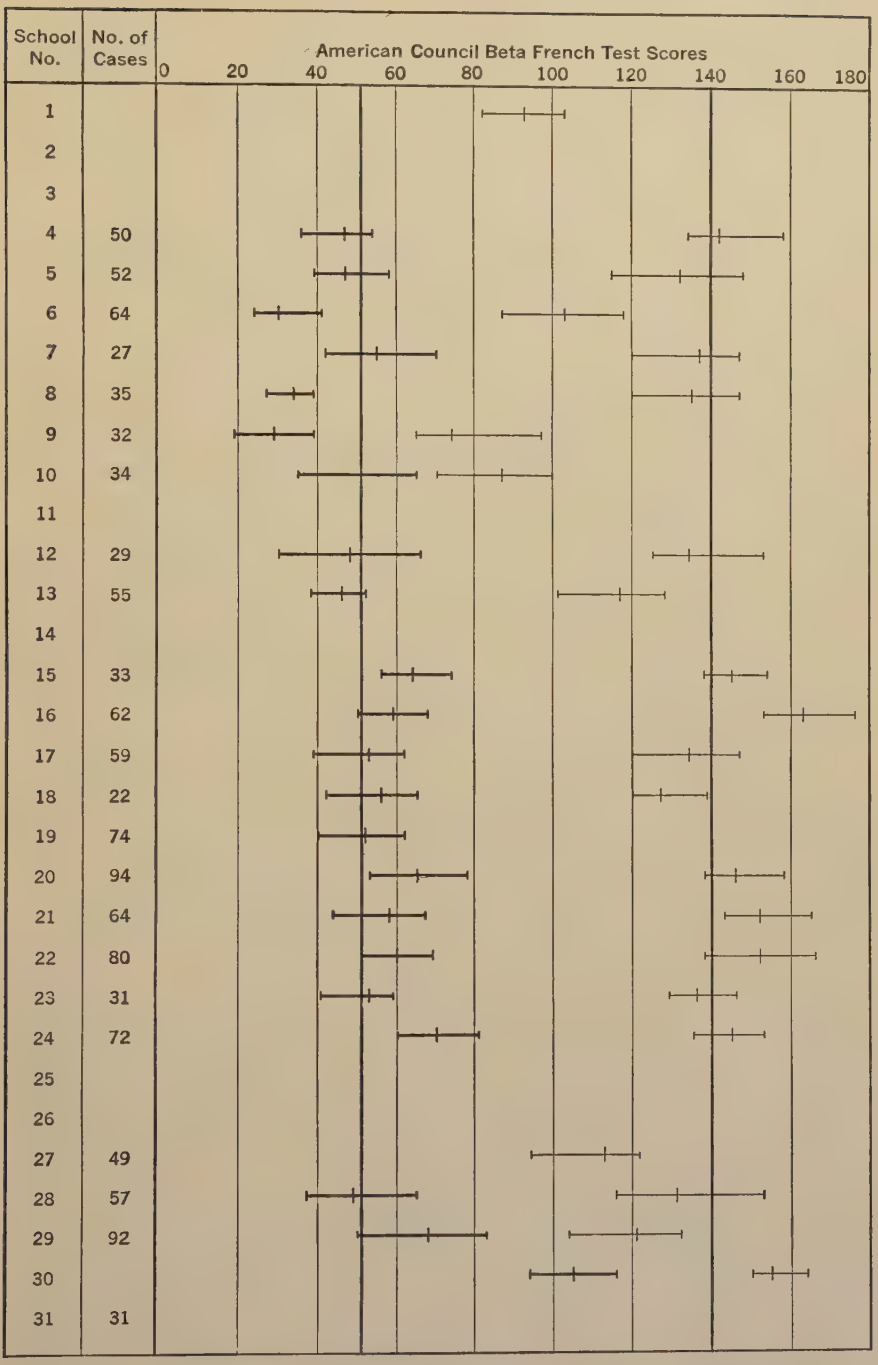


CHART 52. — Medians and interquartile ranges of American Council Beta French Test scores of identical groups of students who took Form A in June, 1925 as RB students and Form B in June, 1926 as RD students. The heavy vertical lines at 51 and 140 indicate the approximate city-wide RB and RD medians, respectively. This chart is similar to and should be read in the same way as Chart 50, *q. v.*

in 1926, or who were in the 9A 1926 class and not in the 8A 1925 class in the same school, were eliminated from consideration, so far as Chart 50 is concerned. This is also true of the 8B–9B and RB–RD groups treated in Charts 51 and 52. With rare exceptions this required the elimination of only a few students. The schools are arranged in the order of the magnitudes of the Form A 1925 8A medians on Charts 50 to 52, inclusive.

Chart 50 is fairly typical of all three of these charts. Comparing the first and last pairs of interquartile ranges on Chart 50, we find a very striking difference. In school 1 the 8A class in June, 1925, had a median score almost 20 points below the city-wide 8A median. During the 1925–1926 session the 52 students in this class became 9A students and progressed at a rate which placed their median in June, 1926, 16 score-points below the 9A city-wide median. Their rate of progress in the second and third semesters of their modern language work was consistent with that in their first semester. In school 31 72 8A students secured a median score in June, 1925, just five or six points below the city-wide 9A median; during the 1925–1926 session these 72 students became 9A students, thus advancing a whole year in *credit* status but increasing their median score by less than three points! The school situation which produced this incredible result is certainly worthy of most careful investigation. Similarly the situation in school 4, which carried a sub-average 8A class of 23 students in one year to a median achievement only slightly below the 9B city-wide average, is worthy of careful study and emulation.

In Charts 51 and 52 the schools are in the same order as in Chart 50. This fact enables the reader to secure a concrete picture of the confusion of standards and variations in efficiency within a single school. For example, in Chart 50 school 3 is below average and in Chart 51 it is distinctly above average. In Chart 50 school 23 started above average and ended below average, and in Chart 51 school 23 started below average and ended above average. Other striking variations in these charts are left to the interest of the reader.

Similar charts have been prepared for corresponding groups of Spanish classes, but their indications are so closely parallel to those of Charts 50, 51, and 52 that, to conserve time and space, they are omitted from the present report.

Summary.—The charts amply confirm four conclusions arrived at in preceding pages.

1. There is little or no relation between the progress of classes composed of the same individual students in 1924–1925 and in 1925–1926.
2. There is little or no relation between progress and degree of homogeneity of classes.
3. The individual classroom situation is more potent in determining the progress of the class than any other influence that we can isolate.

4. The heterogeneity of individual classroom situations, displayed by these charts, superimposed on the heterogeneity of initial class averages, plus the fact of the imperfect reliability of the tests, is more than enough to mask whatever relation there might be under ideal conditions between homogeneity and achievement, on the one hand, and between initial and final achievement of individual students and of classes, on the other. In other words these charts are more than enough to account for the zero correlations reported above on page 327 and for the small magnitude of the correlations in Tables 48 and 49. Indeed, that there should be any positive correlation between the 1925 and 1926 achievements of individual students under the conditions disclosed by these charts, is a very striking manifestation of the strength of the tendency toward constancy of achievement rates in individual students.

The importance of the question of the constancy of achievement rates in individual students is self-evident, because guidance is in effect prediction, and we have no way of judging the future but by the past of a student. So long as the real potentialities of individual students and of classes are masked by being made the sports of chance in the chaos of school and classroom situations uncovered by these charts, the effective educational guidance of students will remain a pious hope. Nothing short of a large program of research with the use of accurate and comparable measurements, and of continuous student-accounting, will avail in this situation.

For example, many important questions on which our data would have shed light have not been considered in these pages. Among these are: What was the fate, during 1925–1926, of students who showed themselves in 1924–1925 to be exceptionally gifted or exceptionally deficient with regard to modern language achievement? What was the relation of progress in 1925–1926 to (*a*) kind of textbooks used, (*b*) variety of textbooks used, (*c*) shifts in textbooks, (*d*) methods of instruction, (*e*) kind of vocabulary presented, and (*f*) to rate of presentation of vocabulary and grammatical materials? What was the relation of sex to amount and rate of achievement in 1925–1926? (The data were especially suited to the study of this question, since we could compare the progress of boys and of girls in terms of equal units, with other factors rendered constant, such as age, initial learning, etc.) What was the relation of age, and of language spoken at home, and of taking other languages, to achievement in modern language work?

While an investigation of these and other similar questions cannot be undertaken in this report, the conclusions that have been arrived at are of far-reaching significance; they will illuminate other diagnostic studies that might be made of present and of future conditions in modern language teaching, and are at once suggestive of immediate corrective measures.

Date Due

BRODART, INC. Cat. No. 23 233 Printed in U.S.A.

Wood, B. D.
author New York experiments with new-type modern language tests
title
LB 3060.26 .W6 G01808

LB 3060.26 .W6
Wood, Benjamin DeKalbe, 1894-
New York experiments with new-type modern language tests :

BISHOP'S UNIVERSITY LIBRARY
LB 3060.26.W6 General
New York experiments with new-ty
3 1843 00225184 7